THE UNIVERSITY
of EDINBURGH

Mapping the *Shh* regulatory landscape

Eve Anderson

A thesis submitted for the degree of Doctor of Philosophy

University of Edinburgh

2014

Declaration

I hereby declare that this thesis has been composed by me for the degree of PhD at the University of Edinburgh. As far as I am aware, this work has not been presented in any previous application for a degree. All work was performed by myself, unless otherwise stated in the text, and all sources of information have been acknowledged in the text.

Eve Anderson
2014

1. Abstract

Sonic hedgehog (Shh) is an important signalling protein expressed extensively in development, throughout tissues of the central nervous system, gut and the posterior of the limb bud. The complicated expression pattern of *Shh* is regulated by a series of long-range enhancers located flanking and dispersed throughout a 1 Mb genomic desert. Disruption of *SHH* as a result of mutations within the gene or its enhancers has been implicated in two developmental conditions. These are Holoproencephaly (HPE3) a common developmental defect of the forebrain and frequently the mid-face in humans, and preaxial polydactyly (PPD), a congenital limb abnormality encompassing a varied phenotype affecting the digits on the anterior side of the hands and feet that has been attributed to misexpression of *Shh*.

In order to investigate the *Shh* regulatory landscape and survey regulatory activity, a transposon-based chromosomal engineering strategy known as the local hopping enhancer detection (LHED) system was employed. Using this method a targeting vector containing a *LacZ* reporter gene as well as *LoxP* sites was inserted within the *Shh* region. The 'hopping' nature of the transposable element was then exploited to scatter it throughout the region. Tetraploid complementation embryos derived entirely from ES cells were generated and examined in order to gain an insight into enhancer activity. The region was found to be in an open conformation over its length and is generally susceptible to all *Shh* enhancers. Genes within the regulatory domain, such as the widely expressed *Rnf32* gene, were found to resist *Shh* enhancer activities by a process of regulatory evasion by the promoter, a mechanism that may be common in large regulatory domains. Finally, at the boundaries of the region *Shh* activity was found to be lost incrementally at a number of genomic positions.

Mouse lines were also generated to look at both enhancer activity and loss of function effects and three deletions of increasing size were generated between *Shh* and the furthest enhancer, the *Zrs*. These in turn, delete firstly a gut and pharyngeal epithelial enhancer, secondly the gut, pharyngeal enhancers as well an oral epithelial enhancer and finally all three epithelial enhancers as well as three forebrain enhancers. Reporter gene expression was found to be lost incrementally from those enhancers deleted without disrupting the rest of the region. Previously unidentified notochord enhancer(s) were found to lie within the region 100-530 kb upstream of *Shh*. Examination of the resultant phenotypes showed that gut and craniofacial defects were found to occur as a result of the loss of enhancers which drive expression within these tissues. Variable phenotypes were found to occur potentially as a result of temporal changes to *Shh* expression or as a result of threshold levels of HH being required for normal development. Other enhancers within the *Shh* region and outwith the deletions were not found to be disrupted by these modifications suggesting the enhancers within the region act independently of each other. The largest deletion resulted in bringing the *Zrs* (which drives Shh limb expression) within 170 kb of the gene, however limb development; was not, found to be affected suggesting distance is not required for *Zrs* function.

Overall, the LHED transposon system has been utilised in order to examine the *Shh* region in more detail, allowing mapping of enhancer function by reporter gene expression and examination of phenotypes generated by deletion of enhancers.

Table of Contents

1. Abstract 2
List of Figures 9
List of Tables 10
Abbreviations 11
2. Introduction 14
2.1 The Hedgehog genes 14
2.2 The Hh signalling pathway 15
2.3 Shh signalling within the central nervous system 20
2.4 Shh signalling within the Gut 23
2.5 Shh signalling within the limb 25
2.6 Loss of *Shh* in development 29
2.7 Misexpression of Shh in development 32
2.8 The *Shh* gene desert 35
2.9 Conservation of *Shh* enhancers 38
2.10 Mutations affecting *Shh* regulation 39
2.11 Neighbouring genes to *Shh* 41
2.12 Long range Regulation 42
2.13 Conservation of Gene deserts 42
2.14 Long range regulation in Development 43
2.14.1 Sox9 44
2.14.2 Pax6 45
2.14.3 Gremlin 45
2.14.4 The HoxD cluster 46
2.14.5 Fgf8 47
2.15 Genomic Landscapes 47
2.15.1 What directs enhancers towards their correct promoter? 47
2.15.2 What prevents enhancer activity? 53
2.17 Aims 58
2.17.1 Investigating the composition of long-range regulation within the *Shh* locus. 58
3. Materials and Methods 60
3.1 Plasmids 60
3.1.1 DNA BAC 60
3.1.2 Plasmids 60

3.1.3 LHED system 60
3.2 DNA spectrophotometry 60
3.3 Agarose Gel electrophoresis 60
3.4 Sequencing 60
3.5 Nucleic acid extraction 61
3.5.1 DNA extraction from mouse tissues 61
3.5.2 DNA extraction from ES cells 61
3.5.3 RNA extraction from mouse tissues and cell cultures 61
3.6 Polymerase Chain Reaction 61
3.6.1 Proofreading PCR 62
3.6.2 Standard PCR 62
3.6.3 Long Range PCR 62
3.6.4 Nested PCR 62
3.7 Single-stranded cDNA synthesis & RT-PCR 63
3.8 Phenol chloroform/chloroform extraction 63
3.9 DNA precipitation 63
3.10 Restriction Enzyme Digest 64
3.11 Ligations 64
3.12 Transformation 64
3.13 Bacterial recombineering 64
3.14 Small-Scale Plasmid Purification 65
3.15 Large-scale Plasmid Purification 65
3.16 Gel Extraction 65
3.17 ES Cell Culture 65
3.17.1 Maintenance of ES Cells 65
3.17.2 E14IVtg2a /G4 ES cell culture 65
3.17.3 Cell Splitting 66
3.17.4 Cryopreservation 66
3.17.5 Gene Targeting in ES Cells 67
3.17.6 Picking Colonies 67
3.17.7 Transient Expression of Cre 67
3.17.8 ES cell transfection 67
3.17.9 Blastocyst Injections 68
3.17.10 Tetraploid complementation assay 68

3.17.11 Retinoic Acid treatment of ES cells 69
3.17.12 Production of Methaphase nuclei 69
3.18 Probe labeling (Biotin) 69
3.19 Fluorescent *in situ* hybridization (FISH) 69
3.20 Histology 70
3.20.1 Whole mount X-gal staining 70
3.20.2 RNA probe synthesis 71
3.20.3 *In situ* hybridisation 71
3.20.4 Haematoxylin and Eosin staining 72
3.20.5 Alkaline phosphatase staining 73
3.21 Microscopy 73
3.21 OPT analysis 73
3.22 Mouse lines 74
4. Results 78
4.1 Introduction 78
4.2 Designing the LHED strategy 80
4.2.1 Designing homology arms for exploration of the Shh genomic region 81
4.2.2 PCR amplification and pBluescript cloning 82
4.2.3 Bacterial recombineering 83
4.3 G4 ES cell targeting 85
4.3.1 PCR analysis of targeted clones 85
4.4 E14 ES cell targeting 87
4.4.1 PCR analysis of E14 targeted ES cells 87
4.5 FISH analysis of correctly targeted ES cells 90
4.6 Inducing transposition in targeted ES cells 91
4.7 Determining 'Hopping' and Reinsertion within LHED_SHH2 Targeted cell Lines ... 94
4.8 Determining 'Hopping' and Reinsertion within lmbr1_LHED Targeted cell Lines ... 95
4.9 Conclusions 97
5. Mapping the *Shh* regulatory region 100
5.1 Introduction 100
5.2 Generation of tetraploid embryos 101
5.3 Reporter activity across the gene desert 101
5.4 Do Insertions vary in their intensity of staining? 103
5.5 Reporter Gene Activity inside Genes 105

5.6 Defining the boundaries of *Shh* enhancer activity 109
5.7 How robust are enhancers within the *Shh* regulatory region?..... 111
5.8 Does the transposon vector carry any genetic information when it jumps?..... 113
5.9 What are the expression patterns of neighbouring genes within the *Shh* region?..... 114
5.9.1 *In situ* hybridisation probe cloning 114
5.9.2 Is the *Zrs* capable of acting on *Mnx1*? 115
5.10 Is there a topological domain surrounding *Shh*?..... 118
5.11 Conclusions..... 120
6. Analysis of LHED mouse lines..... 123
6.1 Introduction..... 123
6.2. Mouse line production 124
6.2.1 Expression analysis..... 124
6.2.2 Phenotype analysis..... 125
6.3 How do deletions within the *Shh* region affect enhancer activity and development? 130
6.3.1 *LacZ* reporter gene analysis of SBLac lines..... 130
6.3.2 LacZ reporter gene analysis of g/2-120 131
6.3.3 Phenotype analysis of the g/2-120 line 133
Summary..... 135
6.3.4 LacZ reporter gene analysis of g/2-67 heterozygote..... 137
6.3.5 Phenotype analysis of the g/2-67 line 138
Summary..... 142
6.3.6 LacZ reporter gene analysis of g/120..... 144
6.3.7 Phenotype analysis within the g/120..... 146
Summary..... 153
6.4 How does the presence of multiple promoters affect enhancer activity?..... 158
6.5 Conclusions..... 164
6.5.1 Expression analysis..... 164
6.5.2 Phenotype analysis..... 167
6.5.3 The effect of multiple promoters on enhancer activity 170
7. Discussion..... 172
7.1 Introduction..... 172
7.1.1 Targeting of an LHED vector into the *Shh* region 172
7.1.2 Remobilisation of the LHED vector 173
7.2 Use of the LHED system *in vitro*..... 174

7.3 Mapping the *Shh* region using LHED 175
7.4 The effect of deletions on reporter activity 178
7.5 The effect of deletions on phenotype 180
7.7 Perspectives and further work 184
8. Appendix 187
8.1 Developing a FISH/ES cell system to establish the chromosomal dynamics at the *Shh* locus 187
8.1.1 Inducing Shh expression within ES cells 187
8.1.2 What are the chromosomal dynamics at the *Shh* locus? 187
8.2 Generating deletions in vitro to examine the chromosomal dynamics at the *Shh* locus 190
8.3 Generation of deletion lines upstream of the ZRS 191
8.3.1 g/3-21 mouse line expression/phenotype analysis 192
8.3.2 g/3-119 mouse line expression/phenotype analysis 192
8.4.1 Designing a targeting vector to 'Knock-out' Lmbr1 193
8.4.2 Vector Cloning 193
8.4.3 ES cell Targeting 194
8.5.4 *Lmbr1* 'Knock-out' Mouse production 196
8.6 Conclusions 196
Location of LHED reinsertions 199
9. Acknowledgments 205
References 206

List of Figures

Figure 2.1: The Hedgehog signalling pathway 19
Figure 2.2: Patterning of the developing limb bud 28
Figure 2.3: Regulation of Shh expression in development 35
Figure 2.4:HoxD Regulatory Landscape..... 51
Figure 2.5: FGF8 Regulatory Landscape 53
Figure 4.1: The LHED transposon system 79
Figure 4.2: LHED targeting into the *Shh* region 82
Figure 4.3: Homology Arm Cloning 84
Figure 4.4: Homologous recombination and PCR design..... 86
Figure 4.5: ES cell targeting 89
Figure 4.6: FISH analysis of correctly targeted ES cell lines 91
Figure 4.7: Mapping reinsertions at the Shh locus in LHED-targeted ES cells..... 95
Figure 5.1: Enhancer activity within the gene desert 102
Figure 5.2: Enhancer activity time course..... 105
Figure 5.3: Enhancer activity at the 5' end of the *Shh* regulatory landscape 108
Figure 5.4: Enhancer activity at the 3' end of the *Shh* regulatory landscape 111
Figure 5.5: OPT analysis of insertions across the *Shh* region..... 112
Figure 5.6: LHED transposition does not result in insertions within the genomic region..... 114
Figure 5.7: Expression of genes neighbouring the *Shh* regulatory landscape..... 117
Figure 5.8: Summary of the topological domains surrounding the Shh regulatory region..... 119
Figure 6.1: Deletions within the *Shh* region..... 124
Figure 6.2: Mouse crosses and resulting pups 129
Figure 6.3: Reporter gene activity in a LHED control line..... 131
Figure 6.4: Reporter gene activity in the heterozygous g/2120 mouse line..... 133
Figure 6.5: Reporter gene activity within sagittal head sections from LHED mouse lines 133
Figure 6.6: Phenotype analysis of the g/2-120 mouse line 136
Figure 6.7: Reporter gene activity in the g/2-67 mouse line..... 138
Figure 6.8: Phenotype analysis of the g/2-67 mouse line 139
Figure 6.9: Histological analysis of the g/2-67 mouse line..... 141
Figure 6.10: Reporter gene activity in the heterozygous g/120 mouse line..... 145
Figure 6.11: Phenotype analysis of the g/120 mouse line at E12.5 147
Figure 6.12:Phenotype analysis of mildly affected g/120 mice at E17.5..... 149
Figure 6.13: Phenotype analysis of moderate/severely affected g/120 mice at E17.5 152
Figure 6.14: The effect of multiple promoters within a genomic locus..... 159
Figure 6.15: Phenotype analysis of SBLac120 homozygous mice at E17.5..... 162

Figure 6.16: How do deletions within the *Shh* region affect reporter gene expression and phenotype? 164
Figure 8.1: Chromosomal dynamics at the *Shh* regulatory locus 189
Figure 8.2*: in vivo* LHED deletions within the *Shh* region 190
Figure 8.3: Cloning of an upstream Lmbr1 targeting vector 195

List of Tables

Table 3.1: Primer Sequence 74
Table 4.1: Inducing hopping with LHED targeted ES cells 92
Table 4.2: Summary of transfections used to optimise transposase activation within LHED target cell lines 93
Table 4.3: Distribution of LHED insertions across the mouse genome and within chromosome 5 96
Table 8.1: Reinsertion sites of the LHED transposon after SB100X mobilzation 199

Abbreviations

Abbreviation	Abbreviated term
AER	Apical ectodermal ridge
AP	Alkaline phosphatase
A-P	Anterior-posterior
BAC	Bacterial artificial chromosome
BLAT	Basic local alignment tool
Bp	Base pairs
BSA	Bovine serum albumin
CD	Campomelic dysplasia
cDNA	Single stranded complementary deoxyribonucleic acid
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CNCCs	Cranial neural crest cells
CNS	Central nervous system
Cos2	*Costal2*
CPPs	Cardiac Pulmonary Progenitors
CRD	Cysteine-rich-domain
DAPI	6-diamidino-2-phenylindole dihydrochloride
Δ	Deletion
DEPC	Diethyl dicarbonate
DIG	Digoxigenin
Dhh	*Desert Hedgehog*
DMEM	Dulbecco's modified Eagles medium
DMSO	Di-methyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
D-V	Dorso-Ventral
E	Embryonic day
ES cell	Embryonic stem cell
EDTA	Ethylenediaminetetraacetic acid
En2	*Engrailed2*
ER	Endoplasmic reticulum
F1	Filial 1
FCS	Foetal calf serum
FEZ	Frontonasal ectodermal zone
FISH	Fluorescent *in situ* hybridisation
FGF	Fibroblast growth factor
Fmn1	*Formin1*
FNP	Frontonasal prominence
FRT	Flippase recognition site
Fu	*Fused*

Fz	*Frizzled*
GCR	Global control region
GFP	Green fluorescent protein
GPCRS	G-protein coupled receptor superfamily
Grem1	*Gremlin*
Hh	*Hedgehog*
Hm	*Hammertoe*
HPE	Holoproencephaly
H&E	Haematoxylin & Eosin
Hx	*Hemimelic extra toes*
Ihh	*Indian Hedgehog*
LCR	Locus control region
LINE	Long Interspersed Elements
LGE	Lateral ganglionic eminence
LHED	Local Hopping Enhancer Detection
LIF	Leukaemia inhibitory factor
Lnp	*Lunapark*
Mb	Mega base pairs
Mge	Medial ganglionic eminence
n	Number
nt	Nucleotide
O/N	Overnight
OPT	Optical projection tomography
PA1	First pharyngeal arch
PBS	Phosphate buffered saline
PBST	PBS Tween-20
PCR	Polymerase chain reaction
PFA	Para-formaldehyde
PPD	Preaxial Polydactyl
Ptc	*Patched (Drosophila)*
Ptc1	*Patched1 (Vertebrates)*
MACS1	Mouse amphibian conserved sequence 1
MFCS4	Mouse fish conserved sequence 4
MRCS1	Mouse reptile conserved sequence 1
NCCs	Neural crest cells
RA	Retinoic acid
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT	Room temperature
RT-PCR	Reverse transcription-PCR
SBE1	Shh brain enhancer 1
SBE2	Shh brain enhancer 2

SBE3	Shh brain enhancer 3
SBE4	Shh brain enhancer 4
SD4	Syndactyly type 4
SLGE	Shh lung gut enhancer
SFPE1	Shh floorplate enhancer 1
SFPE2	Shh floorplate enhancer 2
Shh	*Sonic Hedgehog*
SINE	Short Interspersed Elements
Smo	*Smoothened*
SSC	Saline sodium citrate
Ssq	*Sasquatch*
SuFu	*Suppressor of Fused*
TBE	Tris/borate/EDTA
TBS	Tris buffered saline
TBST	Tris buffered saline Tween-20
TNF	Tumour necrosis factor
TPTPS	Triphalangeal thumb polysyndactyly
U	Units of specific enzyme activity
Wg	*Wingless*
wt	Wild type
YAC	Yeast artificial chromosome
Zli	Zona limitans intrathalamica
ZPA	Zone of polarizing activity
Zrs	ZPA regulatory sequence

2. Introduction

2.1 The Hedgehog genes

The hedgehog gene (*Hh*) was first discovered by (Nusslein-Volhard and Wieschaus 1980) within a mutation screen of *Drosophila* embryos. *Hh* has been shown to encode a secreted protein which is involved in patterning the *Drosophila* embryo. It is confined to engrailed-expressing cells and serves to pattern the posterior compartment of each segment primordium (Lee et al. 1992). Loss of function mutations result in a phenotype whereby the denticles found on the larval cuticle resembled the spines of a hedgehog, thus giving the gene responsible its name (Nusslein-Volhard and Wieschaus 1980).

Vertebrate hedgehog genes were first identified in mice, fish and chickens via a cross-species screen (Echelard et al. 1993; Krauss et al. 1993; Riddle et al. 1993). Unlike the fly where only a single *Hh* gene was identified, Within the mouse three *Hh* genes are known- Desert Hedgehog (*Dhh*) which is most closely related to the *Drosophila Hh* gene; Indian hedgehog (*Ihh*) and Sonic Hedgehog (*Shh*) (Echelard et al. 1993).

In mice, *Shh* has been shown to be essential for patterning the early embryo. *Shh* is expressed at Hensen's node, throughout the notochord, the floor plate of the neural tube, and the early gut endoderm, and in the posterior of the limb buds. It thus plays a key role in patterning of the ventral neural tube (Echelard et al. 1993), the anterior posterior limb axis (Riddle et al. 1993), the epithelial morphogenesis of the oral cavity to hindgut (Chuong et al. 2000) and the ventral somites (Johnson et al. 1994).

Ihh is expressed at two main sites within the embryo; the endoderm of the mid-gut and lung and the cartilage of the developing long bones in the limb. Within

the limb, *Ihh* is normally expressed within the growth plate where it is responsible for regulation of chondrocyte proliferation and differentiation. *Ihh* is not found to be expressed early in development when *Shh* is expressed in the posterior mesenchyme, suggesting that *Ihh* has a distinct role from Shh. Despite this, Ihh has biological properties similar to that of Shh, including the ability to regulate the conserved targets *Patch1* (*Ptc1*) and *Gli* (Vortkamp et al. 1996).

In mice, *Dhh* has been found to play a role in male sexual development and in the formation of the perineurium, which protects the nerve fibres of the peripheral nervous system. *Dhh* has been found to be expressed within foetal testes at E11.5, with no transcripts being detected in foetal ovaries (Bitgood et al. 1996).

2.2 The Hh signalling pathway

The hedgehog signal transduction pathway is an evolutionary conserved signalling cascade essential for the proper patterning and development of tissues in metazoan organisms.

The Hh protein is initially synthesised as a 45 KDa precursor protein, which undergoes an autoproteolytic cleavage via an intein-like mechanism. The C-terminal domain of Hh, which resembles a self-splicing intein domain, catalyses both its own removal and replacement by cholesterol (Hall et al. 1995). Two fragments are generated; a 25 KDa non-functional C-terminal fragment and a 19 KDa N-terminal fragment, which is covalently linked with cholesterol. The N-terminal domain is further modified within the endoplasmic reticulum by a protein known as skinny hedgehog which palymitoylates a Cys residue near its N-terminus. Once both lipid modifications are performed the protein is packaged and released in lipid-associated particles (Hall et al. 1995).

The Hh ligand functions as a morphogen that has been implicated in both short and long range signalling. Within the vertebrate, Shh is expressed within the notochord where it acts locally at high concentrations to induce floor plate specification (Roelink et al. 1995). Whereas, at larger distances over ~200 µm, and

lower concentrations it is found to contribute to the specification of motor neurons and the induction of Ptc1 (Ericson et al. 1996; Marigo and Tabin 1996; Ericson et al. 1997). Within the *Drosophila* embryo Hh acts locally to regulate Wingless (Wg) within cells adjacent to those expressing the Hh protein (Ingham 1993). A concentration dependent gradient of Hh protein is necessary for patterning the adult fly abdomen (Struhl et al. 1997). Hh has also been shown to act at long-range to pattern the anterior posterior axis of the *Drosophila* wing where is acts over distances of 8-10 cell diameters to induced expression of both Dpp and Ptc (Mullor et al. 1997).

It was originally believed that the molecular mechanism of Hh transduction in responding cells differs significantly between *Drosophila* and mammals (Huangfu and Anderson 2006; Varjosalo et al. 2006). However, more recent evidence suggests these pathways, although somewhat different, have many similarities.

In responding cells, the Hh protein binds to Patched (Ptc) a 12 pass transmembrane protein which functions as its core receptor and works in combination with co-receptors such as Ihog/Cdo (Marigo, Davey, et al. 1996; Stone et al. 1996; Tenzen et al. 2006; Zhang et al. 2006). In the absence of Hh, Ptc represses the activity of the 7-pass transmembrane protein Smoothened (Smo), a member of the G-protein coupled receptor superfamily (GPCRS), which has been shown to strongly resemble the Frizzled (Fz) family of proteins (Nusse 2003). Both Smo and Fz have an N-terminal extracellular cysteine-rich domain (CRD). The mechanism by which Ptc represses Smo activity still remains unclear. Early reports suggested that Ptc represses Smo by directly binding to it via the CRD domain much like Fz and Wnt family ligands, however this was later shown to be an artefact of the over-expression experiments used and that actually Smo inhibition is achieved by sub-stoichiometic amounts of Ptc (Stone et al. 1996). The Ptc protein has similar structure to cholesterol trafficking regulator Niemann-Pick C1 (Npc1) protein. Like Npc1, Ptc contains a sterol-sensing domain suggesting it could play a role in transporting hydrophobic molecules that regulate Smo activity (Huangfu and Anderson 2005). Smo signalling activity has been shown to be modified by sterol derivatives (Eaton 2008). A more recent study suggests that Smo activation involves

a conformational rearrangement modulated by lipidic molecules acting at multiple distinct sites within its extracellular and heptahelical domains suggesting Smo activity is determined by multiple physiological inputs (Myers et al. 2013). Ptc might therefore function by regulating the availability of such molecules, with its similarity to Npc1 suggesting it could promote transport of small lipophilic molecules (Eaton 2008).

Once Smo inhibition is released it forms an active conformation. In *Drosophila* this is communicated to the cytosol through a signalling complex comprising the Costal 2 (Cos2), Fused (Fu) and Suppressor of Fused proteins (SuFu). In the absence of Hh signalling, Cos2 forms a complex with Fu, SuFu and the Ci zinc-finger transcription factor that both promotes cleavage of the full-length Ci to its repressor form and prevents entry of Ci to the nucleus (Ruel et al. 2003). In the presence of Hh a complex is formed between Smo and Cos2, which prevents inhibition of Ci by Cos2 complex generating the activator form of Ci (CiA) which can then move to the nucleus with the help of Fu to activate target genes (Ruel et al. 2003) (Figure 2.1A).

Two mammalian Cos2-like proteins exist, Kif7 and Kif27 which share ~23% and ~38% sequence homology respectively (Katoh and Katoh 2004, 2004). Three homologs of *Ci* are found within vertebrates; *Gli1*, *Gli2* and *Gli3* (Jia et al. 2002; Jia et al. 2005; Bhatia et al. 2006).

Phosphorylation plays an important role in regulating the Hh pathway. When Smo is phosphorylated by PKA, Cos2 is released and a recognition signal for Slimb is established leading to the degradation of the Ci protein by the proteasome into Ci-R. In vertebrates phosphorylation of Gli3 allows its recognition by β-Trcp, resulting in the formation of the repressor molecule (Gli3-R). Activation by Shh relieves the inhibition of Smo by Ptc. Smo becomes phosphorylated by Grk2, binds to β-Arrestin and Kif3a, and is trafficked to the cilium. This relieves the inhibitory effect of Sufu and allows the full-length Gli3 (Gli3-A) to translocate to the nucleus and activate target genes (Jia et al. 2002; Jia et al. 2005; Bhatia et al. 2006).

Hh signal transduction in mammals utilises the primary cilium, an evolutionarily conserved microtubule-based organelle analogous to the flagella, comprising a central microtubule structure known as the axoneme which protrudes from the surface of the cell. This is the major difference between mammals and *Drosophila* with regards the Hh signalling pathway. In *Drosophila* only a few cell types are ciliated and thus they do not appear to require cilia for Hh signalling (Han et al. 2003; Sarpal et al. 2003).

The assembly and disassembly of the cilium is mediated by intraflagellar transport (IFT) proteins and their associated Kinesin II (Kif 3 family), mice mutated in genes associated with these proteins such as *Kif3a* and *Ift88* have been shown to lose Gli repressor and activator function (Rosenbaum and Witman 2002; Huangfu and Anderson 2005). Smo, Ptc1, Gli1, Gli2, Gli3 and Sufu are all found to localise to the primary cilium (Chen et al. 2009). Binding of Hh to Ptc causes removal of Ptc1 from the cilium and translocation of Smo in a Kif3a and β-arrestin dependent manner. In the absence of the Hh ligand Ptc1 inhibits Smo resulting in it being held within intracellular vesicles within the cell, with Gli3 being kept in the primary cilium in a complex with Kif7 and Sufu (Corbit et al. 2005; Rohatgi et al. 2007) (Figure 2.1B).

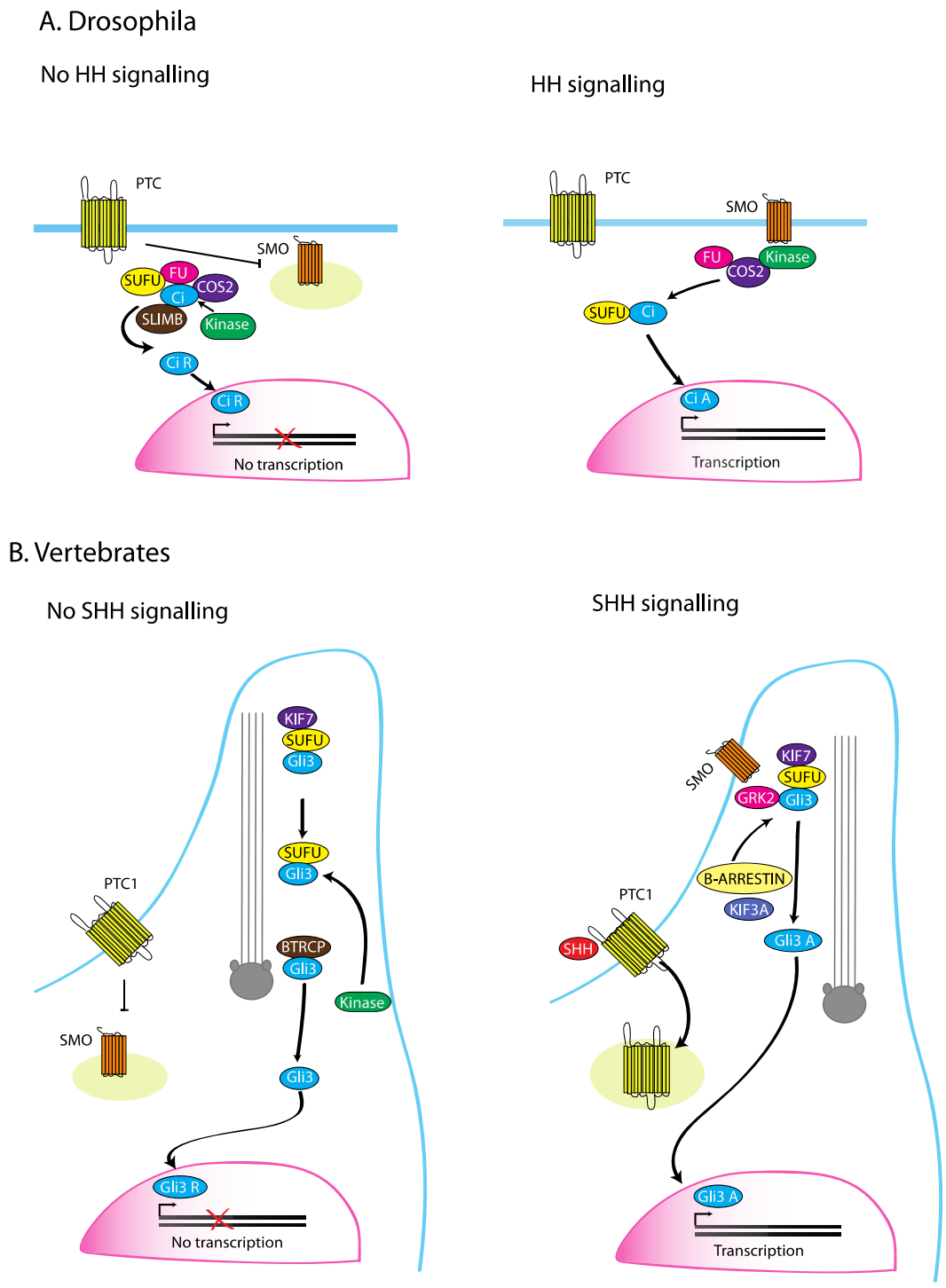


Figure 2.1: The Hedgehog signalling pathway

A. The Drosophila Hh signalling pathway in the absence (left) or presence (right) of Hh. In the absence of ligand, Patched (Ptc) inhibits Smoothened (Smo), which is held in intracellular vesicles. A complex of proteins, including cubitus interruptus (Ci), Costal2 (Cos2) and several kinases, is established. Phosphorylation of CI establishes recognition signals for Slimb leading to partial degradation of Ci by the proteasome and formation of the repressor form (Ci-R). Ci-R then translocates to the nucleus, where it represses transcription of HH targets. Binding of secreted Hh to Ptc, blocks Ptc activity and releases Smo from inhibition.

Smo moves to the plasma membrane, where phosphorylation allows interaction with Cos2. Subsequent phosphorylation of Cos2 by Fu leads to release of unphosphorylated, full-length Ci, which can translocate to the nucleus where it promotes transcriptional activation.

B. The vertebrate Hh signalling pathway. In the absence of Shh ligand, Ptc1 inhibits Smo, which is held in intracellular vesicles. Gli3 is kept in the primary cilium in a complex with Kif7 and Sufu. Phosphorylation of Gli3 by kinases allows its recognition by b-TRCP and leads to partial degradation by the proteasome, resulting in the formation of the repressor molecule. Activation by Shh relieves the inhibition of Smo by Ptc1. Smo becomes phosphorylated by Grk2, binds to β-Arrestin and Kif3A, and is trafficked to the cilium. This relieves the inhibitory effect of Sufu and allows the full-length Gli3 to translocate to the nucleus and activate target genes.

2.3 Shh signalling within the central nervous system

Shh acts as a morphogen during development of the early vertebrate ventral neural tube. Later in the dorsal brain it acts as a mitogen on the progenitors of the cerebellum, tectum, neocortex and hippocampus (Ruiz i Altaba et al. 2002).

Dorosoventral specification of the neural tube is hypothesised to occur via two methods. The general consensus suggests Shh is secreted from the notochord (a mesodermal rod which underlies the neural tube) inducing differentiation of the floorplate (a group of specialised cells located at the ventral midline of the vertebrate neural tube). The floorplate expresses *Shh* in response to the notochord signal (Echelard et al. 1993). This is founded in studies within chick which suggests that the floorplate is unable to develop after removal of the notochord (Placzek et al. 2000). Alternatively, it has been suggested that the floorplate, notochord and dorsal endoderm share a common origin in the Henson's node, thus all represent sources of *Shh* (Teillet et al. 1998).

Shh gradient signalling establishes distinct progenitor domains via the regulation of a group of homeodomain proteins including members of the Pax, NKX, Dbx and Irx families. These proteins are divided into two classes; class I which are synthesised in the presence of Shh and class II which are repressed by Shh. By either repression or activation of these targets Shh expression leads to the definition of five neural progenitor domains within the ventral neural tube. Subsequently, mutual repressive interactions among these factors result in the formation of boundaries (Jessell 2000). Gli1, Gli2 and Gli3 are expressed in gradients within the floor plate

predominantly within ventral, intermediate and dorsal regions, respectively and act as targets of Shh signalling (Ruiz i Altaba, Palma, and Dahmane 2002). *Shh* null mice do not develop floorplate or most ventral cell types suggesting Shh is both necessary and sufficient for inducing ventral-neural cell types (Ruiz i Altaba, Palma, and Dahmane 2002). Loss-of-function *Smo* mutations lead to a more dorsally extended failure in the specification of ventral cell types (Wijgerde et al. 2002). This difference is attributed to the contribution of Ihh from the underlying gut endoderm.

The notochord source of Shh also extends into the brain where it is required for ventrally patterning a number of cell types within the hindbrain and midbrain. Whereas, Shh emanating from the prechordal plate is required for ventral forebrain induction (Fuccillo et al. 2004). Shh is also a target of its own expression, as its activation in the ventral midline of the CNS is dependent on high levels of Shh signalling from the underlying axial mesoderm (Echelard et al. 1993).

As well as having a role as a morphogen, Shh has also been implicated as having a mitotic and antiapoptotic role within the CNS. Experiments during the 1950s showed that in the embryonic chick the brain collapses if the notochord and anterior hindbrain are separated from the neuroepithelium. This is attributed to the notochord secreting a trophic factor important for the expansion of the brain vesicles (Kaellen 1965). More recently, it has been shown that when the notochord is transiently displaced from the midbrain floor plate during development, the brain vesicles exhibit abnormal folding and collapse, as a result of decreased proliferation and increased apoptosis within the midbrain. This has been attributed to decreased SHH levels, as a similar effect can also be induced via the addition of cyclopamine (an inhibitor of *Shh*) or rescued by implanting Shh secreting cells (Britto et al. 2002).

The brain develops from the anterior region of the neural tube and is divided into three primary vesicles; the hindbrain vesicle, midbrain vesicle and forebrain vesicle. The forebrain becomes divided into the diencephalon caudally and the telencephalon rostrally. The telencephalon subsequently differentiates into the olfactory bulbs anteriorly, the cerebral cortex dorsally and the basal ganglia ventrally. From E11 onwards, distinct telencephalic progenitor zones become morphologically apparent. Two structures are formed in the ventral telencephalon;

the lateral ganglionic eminence (LGE) a precursor to the striatum and the medial ganglionic eminence (MGE) (Zaki et al. 2003).

In *Shh* null mice, forebrain development is severely affected. The telencephalon is hypoplastic, uninvaginated and the ganglionic eminences are not morphologically identifiable. Expression of ventral telencephalon genes such as *Nkx 2.1, Ihx6* and *Gsh1* is absent (Pabst et al. 2000; Corbin et al. 2001; Rallu et al. 2002) However, in less severely affected *Shh* null embryos, a small ventral telencephalic domain continues to express genes such as *Gsh2, Mash1* and *Dlx2* (Rallu et al. 2002; Corbin et al. 2003). It is possible that other Hh homologues can pattern the telencephalon or can partially compensate for the absence of Shh. In support of this possibility, mice mutant for both *Shh* and *Ihh* appear to lack all ventral character throughout the CNS (Zhang et al. 2001).

Sources outside the telencephalon also influence its patterning. SHH from the Hensen's node is crucial for specification of the MGE (Echelard et al. 1993; Epstein et al. 1999; Gunhaga et al. 2000). Whereas SHH signalling from the prechordal plate could also be required for specification of the ventral telencephalon (Kiecker and Niehrs 2001).

Conditional gene ablation of *Smo* or *Shh* reveal different telencephalic phenotypes depending of the timing of gene ablation ranging from normal telencephalon development although with reduced olfactory bulb size, to no ventral telencephalon. Early HH signalling between E7.5 and E10.5 is therefore believed to be critical for setting up ventral telencephalic progenitor domains. However, it is not required after E12.5 (Machold et al. 2003; Fuccillo et al. 2004).

The forebrain and epithelia of the facial primordial originate from the same ectoderm. Neural cell fate is thought to be a result of the presence of Bmp antagonists and Fgf signalling (Schuurmans and Guillemot 2002; Meulemans and Bronner-Fraser 2004; Wilson and Houart 2004). The ventral forebrain is also in contact with the mesenchyme of the fronto-nasal process allowing communication between tissues. Cranial neural crest cells which give rise to the facial mesenchyme originate from specific axial positions along the dorsal neural tube (Le Douarin et al.

2004). Chick embryos exposed to cyclopamine at various developmental time points produce a variety of facial malformations the severity of which correlate with the temporal and spatial inhibition of Shh. If Shh signalling is disturbed during gastrulation severe malformations of the forebrain and face occur whereas at later time points a grossly morphologically normal telencephalon results however with facial dysmorphologies (Chiang et al. 1996; Cordero et al. 2004; Marcucio et al. 2005).

2.4 Shh signalling within the Gut

The gut begins to develop within the mouse embryo from E8.0 when the anterior and posterior endoderm invaginates forming foregut and hindgut pockets, which then extend and fuse to form the gut tube. From this tube the digestive system comprising the oesophagus, stomach, small intestine and colon and respiratory systems arise during development (Ramalho-Santos et al. 2000).

Within the developing gut Hh proteins are secreted from within the epithelia and then function within the endoderm on targets such as Ptc1, Smo and Gli (Ramalho-Santos, Melton, and McMahon 2000).

Expression is firstly detected within the hindgut and foregut pockets and then extends towards the mid-gut during gut closure. Broad expression is found within the gut tube until E10.5 and then by stages E11.5-14.5 Shh is down-regulated within the hind stomach, jejunum and ileum. Ihh, in contrast is found to be expressed within the hind stomach to anus (Bitgood and McMahon 1995). At later stages (E18.5) both Shh and Ihh were found to be expressed within the stomach epithelium, the small intestine and the colon. Targets of Hh such as Ptc1, Gli1 and Bmp4 are found to be expressed in the mesenchyme throughout the gastrointestinal tract. *Shh*, *Ptc1* and *Ptc2* transcripts have also been detected within the adult stomach (Ramalho-Santos, Melton, and McMahon 2000).

Both *Shh* and *Ihh* mutants have gastrointestinal defects including a smaller gastrointestinal tract, malrotation of the gut as a result of reduction in smooth muscle

thickness and annular pancreas. Histologically, *Shh* mutants have been shown to have an overgrown stomach epithelium, overgrown villi within the duodenum, imperforate anus and intestinal transformation of the stomach (Ramalho-Santos, Melton, and McMahon 2000).

The embryonic mouse lung begins developing at E9.5, following the invagination of the ventral foregut into a pair of endothelial buds. These buds extend and undergo dichotomous branching between stages E9.5-16.5 in order to form the conducting airways and terminal acini. By E17.0 the terminal buds have dilated in order to form sac-like structures which then go on to form the future alveoli (Ten Have-Opbroek 1991). *Shh* is required for functional lung formation, with transcripts being detected throughout the lung epithelium from E11.5 and this expression continues until birth and peaks just before parturition. Targets of Shh including Ptc1 and Smo are found within the mesenchyme of the developing buds adjacent to Shh expressing epithelial cells (Miller et al. 2001). *Shh* null mice lose lung symmetry as a result of branching defects resulting in a dramatic lung phenotype. The resulting lung lobes fail to develop a network of air sacs, and the oesophagus and trachea fail to divide. Mesenchymal proliferation is severely reduced and various Shh targets are found to be down-regulated (Pepicelli et al. 1998).

During embryonic development, the heart and lung develop in parallel to form the cardiopulmonary circulation system when the lung endoderm protrudes into the cardiac mesoderm. Multi-potent cardiopulmonary mesoderm progenitors (CPPs) are a population of cells that have been identified within the posterior pole of the heart and are marked by expression of genes such as *Wnt2*, *Gli1* and *Isl1*. This population of cells has been shown to arise from cardiac progenitors before lung development begins. CPPs go on to generate the mesodermal lineages within the cardiac inflow tract smooth muscles, proximal and vascular endothelium and pericyte like cells. The regulation of these cells is controlled by expression of Shh from the foregut endoderm which is required for connection of the pulmonary vasculature to the heart (Peng et al. 2013).

Shh expression is observed in the pharyngeal epithelial lining and in the anterior ectoderm-derived epithelial lining from around E11.5-12.5 and is implicated

in the development of the oral cavity, the tooth and the tongue (Cobourne et al. 2004). Shh has been recently shown to be involved in patterning of the fungiform papilla, a complex set of tissues and cell types which sit within the epithelium of the tongue and includes the taste buds. Shh signalling centres have been found to be located within the fungiform placode and are required for taste bud development and maintenance, most likely via stage-specific autocrine and/or paracrine mechanisms and by engaging epithelia-mesenchymal or connective tissue interactions (Liu et al. 2013).

2.5 Shh signalling within the limb

The zone of polarizing activity (ZPA) is a small region of sub apical mesenchyme cells found located at the posterior margin of the early limb bud. A series of experiments in which the ZPA was grafted onto different positions along the anterior-posterior axis of the limb bud led to the belief that the identity of a digit is dependent on the distance from the polarizing region (Wolpert 1969; Tickle et al. 1975). A model was therefore proposed in which the ZPA produces a diffusible morphogen that sets up a concentration gradient across the limb bud and provides cells with positional information.

Shh was shown to be expressed in the ZPA of the chick wing bud, and grafts of transfected chicken embryo fibroblasts expressing Shh were shown to have polarizing activity and were capable of inducing digit formation at the anterior of the limb (Riddle et al. 1993). Shh RNA has been detected ectopically at the anterior as well as at the posterior of the limb bud in mouse mutants with additional pre-axial digits as found in the Sasquatch (*Ssq*) mouse mutant (Sharpe et al. 1999).

Shh was originally believed to pattern the limb bud by forming a gradient across the posterior half of the limb (Gritli-Linde et al. 2001). In many cases the expression domains of *Shh* and *Ptc1* are observed in adjacent embryonic tissues while in other cases they are transiently co-expressed in the same regions which include the neural tube (Echelard et al. 1993), the sclerotome (Fan and Tessier-Lavigne 1994; Johnson et al. 1994), the visceral mesoderm (Roberts et al. 1995), as well as the limb bud

where Ptc is found to be expressed in cells in the posterior 2/3 of the limb bud in response to Shh signalling (Riddle et al. 1993). The timing of Ptc1 induction in these tissues correlated with the timing of *Shh* signalling. In addition, *Ptc1* is ectopically induced in response to Shh misexpression (Marigo, Scott, et al. 1996). Binding of Shh to Ptc is believed to limit the spread of the Shh protein. Hedgehog interacting protein (Hhip1) is up-regulated in response to Shh, and also binds to Shh to limit its range. Both targets of Shh therefore are believed to help create a gradient of Shh expression (Chuang and McMahon 1999). Cdo, Boc and Gas1 are negatively regulated by Shh and are believed to enhance *Shh* signalling. They are expressed in overlapping patterns in the anterior of the limb bud with Boc1 and Gas1 playing roles in facilitating long range Shh signalling (Allen et al. 2007; Allen et al. 2011).

Shh signalling is also mediated through the mutual antagonism of Hand2, a member of the basic helix-loop-helix family of DNA binding proteins and Gli3, a zinc finger containing DNA-binding protein. Just after limb budding, *Hand2* is expressed within the posterior of the limb bud. Hand2 then pre-disposes the expression of Shh either directly or through activation of the 5' *HoxD* complex within the ZPA, leading to determination of the A-P axis (te Welscher et al. 2002). A conditional limb mutation for *Hand2* results in the loss of *Shh* expression and a phenotype resembling the *Shh* null limb (Galli et al. 2010). Once SHH is activated it maintains Hand2 and HoxD expression while regulating members of the Gli family of proteins by proteolytic processing. The function of Shh signalling in digit patterning is to relieve Gli3 repression in cells in the posterior half of the limb bud (Figure 2.2 A). In addition graded Shh signalling in the region is translated into a gradient in the ratio of Gli3A/Gli3R protein present, which results in cells responding in a dose-dependent manner to specify digit identity (Litingtung et al. 2002; te Welscher et al. 2002; Wang et al. 2007). In mammals Gli processing takes place in primary cilia (as discussed above), polydactyly is therefore often found in human patients exhibiting ciliopathies (Quinlan et al. 2008).

Shh is inactive, but primed for expression in the anterior margin of the limb (Amano et al. 2009) and under certain genetic circumstances can be ectopically

expressed resulting in the generation of limb defects such as preaxial polydactyly (Anderson et al. 2012). Several genes have been identified by genetic analysis as being responsible for repressing the expression in the anterior domain. *Alx4* is specifically expressed in the anterior margin of the limb and when deleted results in disrupted anterioposterior polarity meaning *Shh* is induced within the anterior of the limb bud (Panman et al. 2005). At the distal edge of the developing limb bud ETS transcription factor genes, *Etv4* and *Etv5* are expressed maintained by Fgf expression in the apical ectodermal ridge (Lettice et al. 2012). Limb buds that are deficient in both Etv4 and Etv5 ectopically express Shh in a mesenchyme domain at the anterior margin of the limb, indicating that their normal role is to restrict Shh expression to the posterior region of the limb (Mao et al., 2009; Zhang et al., 2009). Although Etv4/5 bind to two sites within the *Zrs*, the binding at one of these sites is sufficient to regulate *Shh* negatively in the anterior domain (Mao et al. 2009; Zhang et al. 2009; Lettice et al. 2012) (Figure 2.2 B).

Members of the ETS transcription factors have been shown to play distinct roles in the spatial pattern of *Shh*. Occupancy at multiple ETS sites, which bind the factors Gabpα and Ets1, regulates the position of the *Shh* expression boundary in the limb, thus defining the ZPA. Multiple binding of Etv4 and Etv5 at the *Zrs*, in contrast, represses ectopic *Shh* expression outside the ZPA. It has been shown that the *Zrs* is open and fully capable of binding to activating factors such as Ets1 in the anterior domain of the limb bud. As a result, new mutant sites are capable of binding ETS factors in both normal and ectopic domains and overriding Etv4/Etv5 repression resulting in preaxial polydactyly (Lettice et al. 2012) (Figure 2.2 C).

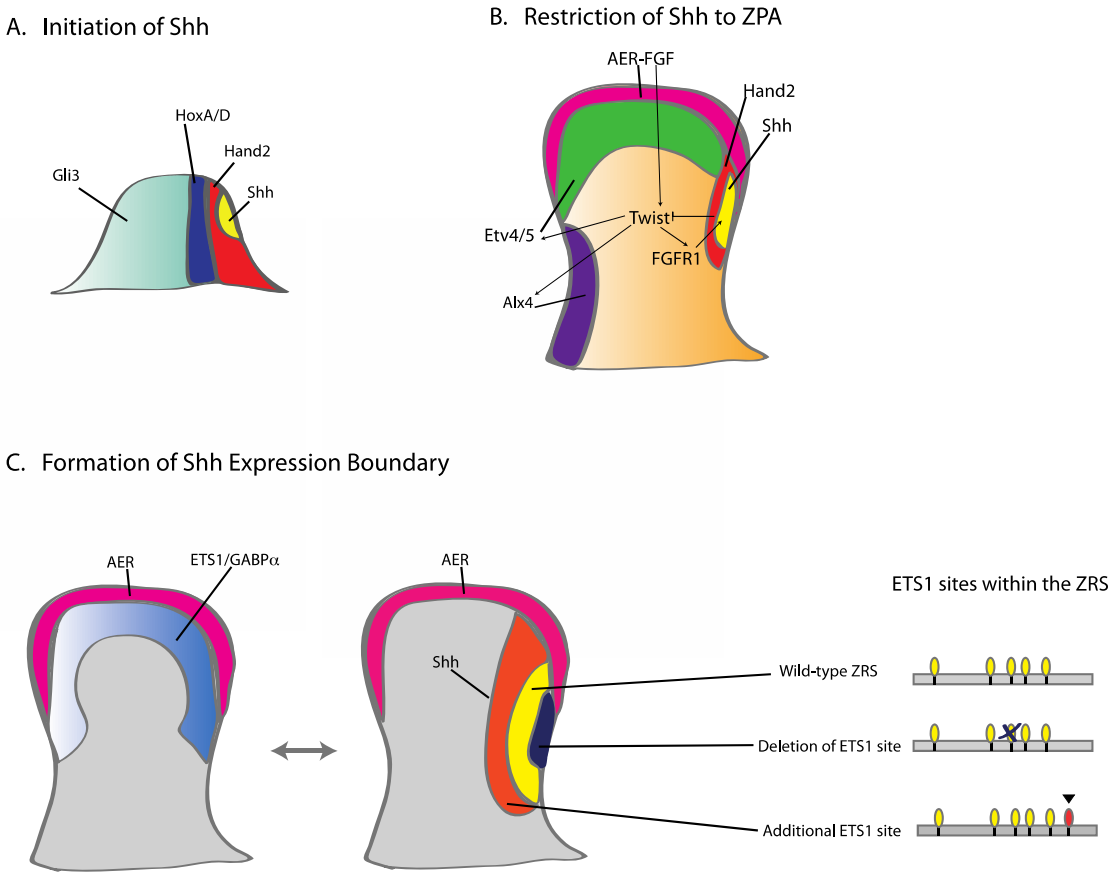


Figure 2.2: Patterning of the developing limb bud

A. Patterning of the early limb bud is defined early in development by the antagonist activity of Gli3 in the anterior and by Hand2 (which down regulates Gli3) in the posterior of the limb bud. Initiation of *Shh* expression is dependent on Hand2 expression and the HoxA/D genes (Hox10-13).

B. Shh expression is repressed in the anterior margin and thus restricted to the posterior margin of the limb. Alx4, Etv4 and Etv5 are involved in this repression. Twist1 expression in the limb up regulates Alx4, Etv4 and Etv5 and in the posterior limb bud the Fgf signalling pathway.

C. The expression boundary of *Shh* is determined by ETS transcription factor binding sites. Within the limb, ETS1/Gabpα is expressed across the limb bud (left). In the wild-type embryo this leads to *Shh* production within the ZPA (right depicted in yellow). Deletion of one of the ETS1 sites causes a reduction in area of Shh expression (right depicted in dark blue) whereas the addition of an Ets1 site results in an increased Shh expression area (right depicted in orange).

The two most posterior digits within the mouse limb are completely derived from the polarizing region. These digits (4 and 5) are both exposed to the same high concentration of Shh in the ZPA but the cells that form digit 4 spend a shorter time within the region before being displaced by growth. This function of Shh is not consistent with a classic morphogen gradient model. It is therefore suggested that

specification of anterior and posterior digits involve different mechanisms; a concentration gradient of Shh which specifies digit 2 and possibly digit 1 and late expansion based differences in duration of Shh signalling which specifies the two posterior digits with digit 4 being specified by a shorter time than digit 5. Finally digit 3 is specified by a combination of Shh concentration and time (Ahn and Joyner 2004; Harfe et al. 2004). However, when Shh expression is reduced at early time points within the limb bud thus reducing cellular expansion, the digits are lost in the order D3, D5, D2, D4 which reflects the reverse order from normal digit primoridia form. Whereas, when Shh is lost at late stages the digits which still form remain morphologically normal despite cell number reduction (Zhu et al 2008). This led to the proposal of a biphasic model in which digit specification takes place very early in mouse limb bud development due to a concentration gradient of Shh expression, and later Shh is required for the proliferation of the specified digit precursors. Shh is therefore hypothesised to act firstly as a morphogen, within the early stages of development, at later stages it then exerts mitogenic activity in order to produce sufficient cell numbers for digit development. These two activities combine to specify digit identity, and as the limb bud expands the position in which these digits form (Zhu et al. 2008). The processing of the Shh protein is believed to be important in determining the range of signalling, this has been demonstrated in mice in which the cholesterol modification of Shh no longer takes place, resulting in middle digit loss (Lewis et al. 2001).

2.6 Loss of *Shh* in development

Shh null mice have been generated by inserting a PGK-Neo cassette within exon 2 of the gene leading to the loss of 97 of the 198 residues and truncation of the protein (Chiang et al. 1996). In contrast to humans where haploinsuffiency at the *Shh* locus occurs, mice heterozygous for the aforementioned mutation were found to be viable and phenotypically normal. However, the disrupted *Shh* gene was lethal when homozygous. Pups are stillborn and expected to have died at or just before birth (Chiang et al. 1996).

The earliest defect exhibited within *Shh* -/- mice is found within the forebrain at E8.5 which corresponds to when Shh is beginning to be expressed within the embryo. At this stage the midline is found to be indistinct and the cephalic folds of the neural tube fused. Additionally, the optic vesicles which are normally separate appear within a single continuous vesicle which protrudes at the midline (Chiang et al. 1996). At the stage of neural tube closure (E9.5) an overall reduction in the size of the brain and neural tube is exhibited. By E11.5 embryos exhibit small size and defects within the presumptive midbrain and forebrain and the externally visible lateral eye structures are absent. By E15.5 embryos exhibit some growth retardation throughout most of the embryo and lack distinct hind limb and forelimbs. Extreme forebrain and craniofacial defects are observed with facial structures such as the eyes, nose and oral structures being mostly unidentifiable (Chiang et al. 1996). By E18.5 histological analysis suggests abnormalities of the heart, lung, kidney and foregut. The limbs are also shown to be completely truncated with the hind limb missing both the tibia and fibula, whereas the forelimb exhibits only a bony extension of the humerus which could represent either a fused ulna/radius or an extended bent humerus. Defects are also found within the ribs with usually only five or six rib cartilages remaining (Chiang et al. 1996).

Within *Shh*-/- embryos defects of the foregut are also exhibited from early stages. At E8.5 within wild type mice outgrowth of the lung, liver and pancreas from the foregut is apparent. However within the mutant, although hepatic buds are formed, development of the lungs is delayed at least half a day. At E9.5 the trachea-oesophageal septum is not established meaning that by E10.5 the trachea and oesophagus are not found as two separate tubes as expected but instead a single trachea-oesophageal tube that connects to the stomach (Litingtung et al. 1998).

By E16.5 the trachea and oesophagus are in close contact, with the mutant trachea surrounded by a cartilaginous structure, a reduction in size of the oesophagus lumen is also found at the level of the heart. Whereas at the level of the lungs the oesophagus is absent and instead a region of thick mucosa resembling the stomach

lining showing continuity with the lung epithelium is found suggesting the oesophagus and lung epithelium have merged (Litingtung et al. 1998).

The lungs are also affected within the mutant; at E9.5 the developing lung is hypoplastic indicating epithelial branching is affected (Litingtung et al., 1998). By E13.5, the asymmetry of the left and right lung lobes is completely absent and the lung appears as a single lobe, which is reduced to a simple sac by E18.5 as a result of enhanced cell death and decreased cell proliferation (Litingtung et al. 1998).

Gastrointestinal defects including gut malrotation, reduced smooth muscle, intestinal transformation of stomach, annular pancreas, duodenal stenosis, and imperforate anus are also observed (Ramalho-Santos, Melton, and McMahon 2000).

In humans the *SHH* gene was identified as a causative locus of HPE, a common developmental defect of the forebrain and frequently the mid-face in humans (Belloni et al. 1996). The condition involves incomplete development and septation of midline structures in the CNS. In its most severe form it results in failure of division of the forebrain into right and left hemisphere, cyclopia, a primitive nasal structure and/or midfacial clefting usually leading to death. Whereas in its mild form signs can include microcephaly, hypertelorism and the presence of a single maxillary incisor as a result of premature fusion of the epithelial dental lamina (Roessler et al. 1996). The condition is highly variable even within families (Roessler et al. 1996).

Shh was chosen as a potential candidate locus for HPE3 as a result of its expression pattern and the phenotype of the *Shh* -/- mouse. Genetic mapping also suggested the *Shh* locus could be involved in the disease and mutations of the gene including missense and nonsense mutations were identified in a number of HPE3 families (Roessler et al. 1996).

In humans, the ventral forebrain is particularly sensitive to the level of *SHH* expression, given that a 50% reduction in its normal levels causes HPE (Roessler et

al. 1996). In mice, loss of both *Shh* alleles is required to generate a similar HPE phenotype (Chiang et al. 1996).

Teratogenic agents such as ethanol and retinoic acid have also been shown to affect SHH. Both these agents have long been recognised to cause central nervous system and craniofacial malformations in humans which are believed to be mediated through the *SHH* pathway (Schneider et al. 2001). In sheep, deformities of the head consisting of distortion or absence of facial bones, cyclopia of the eyes, micropthalmia and anopthalmia with fusion of the cerebral hemispheres and absence or displacement of the pituitary and hydrocephalus as well as, on occasion, limb defects have been shown to be linked to ingestion of *Veratrum californicum* (wild corn lily) by pregnant ewes (Binns et al. 1959). The teratogenic steroidal alkaloid, cyclopamine, was purified from the plant and has been shown to act by inhibiting the cellular response to the Shh signal (Cooper et al. 1998; Incardona et al. 1998). Cyclopamine has also been shown to have a similar effect when administered to chick embryos (Incardona et al. 1998).

2.7 Misexpression of Shh in development

Misexpression of *SHH* is also a major cause of the human congenital limb abnormality PPD (OMIM: 174500); a condition encompassing a varied phenotype affecting the digits on the anterior side of the hands and feet. Patients with preaxial polydactyly exhibit phenotypes ranging from mild triphalangeal thumb to severe duplications of the digits and tibial aplasia (Lettice and Hill 2005). These mutations are dominantly inherited and highly penetrant, and were originally found to map to the 7q36 region within humans (Heus et al. 1999). In mice the dominant mouse mutations *Hemimelic extra toes* (*Hx*) and *Hammertoe* (*Hm*) exhibit limb defects which are analogous to the PPD defects, with the former exhibiting PPD and radial and tibial hemimelia and the latter webbing between the digits (Clark et al 2000). These mutations were shown to map to a homologous region to PPD within humans (Heus et al. 1999; Clark et al. 2000). Originally the PPD phenotype was attributed to mutations within the *Lmbr1* gene as it was present within the mapped locus and was

expressed within the developing limbs at the stage that *Hx* and *Hm* mutations arise (Clark, Marker, and Kingsley 2000; Clark et al. 2001), however comprehensive analysis of the *LMBR1* coding region from cell lines of PPD mutations found no pathogenic mutations of *LMBR1* (Heus et al. 1999; Clark et al. 2001). Mice in which one copy of *Lmbr1* has been deleted exhibited wrist defects in 25% of instances (Clark et al. 2001). However human patients in which one copy of *LMBR1* was lacking were shown to have no limb defects. A mouse was therefore generated in which exon 1 including the translational start site and potentially the transcriptional start site was deleted (Clark et al. 2001). Within these mice around 7% of *Lmbr1* transcripts were still found to be produced (Clark et al. 2001). Histological samples from mice suggested no defects were present within the organs of the liver, kidney, spleen, testes, epididymis or seminal vesicle. There was a low incidence of limb defects; however, in contrast to the *Hx* mutant digit loss rather than digit gain was exhibited (Clark et al. 2001). It was therefore postulated that a gain of function rather than deletion mutation within *Lmbr1* was responsible for the PPD phenotype (Clark et al. 2001).

The *Shh* gene was also found to be physically linked to the PPD phenotype; it was therefore proposed that PPD was in fact occurring as a result of mutations to a long-range regulator of *Shh* (Chang et al. 1994; Sharpe et al. 1999). This hypothesis was made more plausible by data from Holoproencephaly patients who had a translocation breakpoint distal to *Shh* thus suggesting that cis-regulatory control elements were required for normal expression of *Shh* during craniofacial development (Roessler et al. 1997).

Ssq is a mouse mutation which arose through a random transgenic insertion, and presents with a semi-dominant phenotype in which extra digits are present on the hind limbs in heterozygotes and the fore and hind limb of homozygotes. The mutation mapped to intron 5 of the *Lmbr1* gene and was found to be as a result of 20 kb duplication (Lettice et al. 2002). A genetic *cis/trans* test was performed in which a mouse cross was devised whereby the *Ssq* allele was located in *cis* to a distinguishable *Shh* null allele and was used to verify that the *Ssq* phenotype was abrogated when the *Ssq* allele was placed in a *cis* position relative to the *Shh* null

allele (Lettice et al. 2003). A similar test was later performed in *Hx* mice which again showed that the phenotype was abrogated in the presence of the *Shh* null allele (Sagai et al. 2004). This suggested that both mouse mutants were exhibiting phenotypes directly as a result of misexpression of *Shh*. Molecular studies were then performed to confirm that an 800 bp highly conserved long-range regulatory element was present within this region. This regulator was designated the ZPA regulatory sequence (*Zrs*) (Lettice et al. 2003; Sagai et al. 2004). The *Hx* mice as well as four unrelated human PPD families were found to have base substitutions within this conserved sequence, with the conserved region also being capable of driving expression of the *LacZ* reporter gene within the posterior of the developing limb bud in a pattern similar to endogenous *Shh* (Lettice et al. 2003). The ZRS therefore comprises a cis-regulatory element which is both necessary and sufficient for driving spatial and temporal expression of *Shh*, functioning as a long-range regulator, operating over a large distance of approximately 1 Mb to regulate *Shh* expression (Lettice et al. 2003; Sagai et al. 2004) (Figure 2.3).

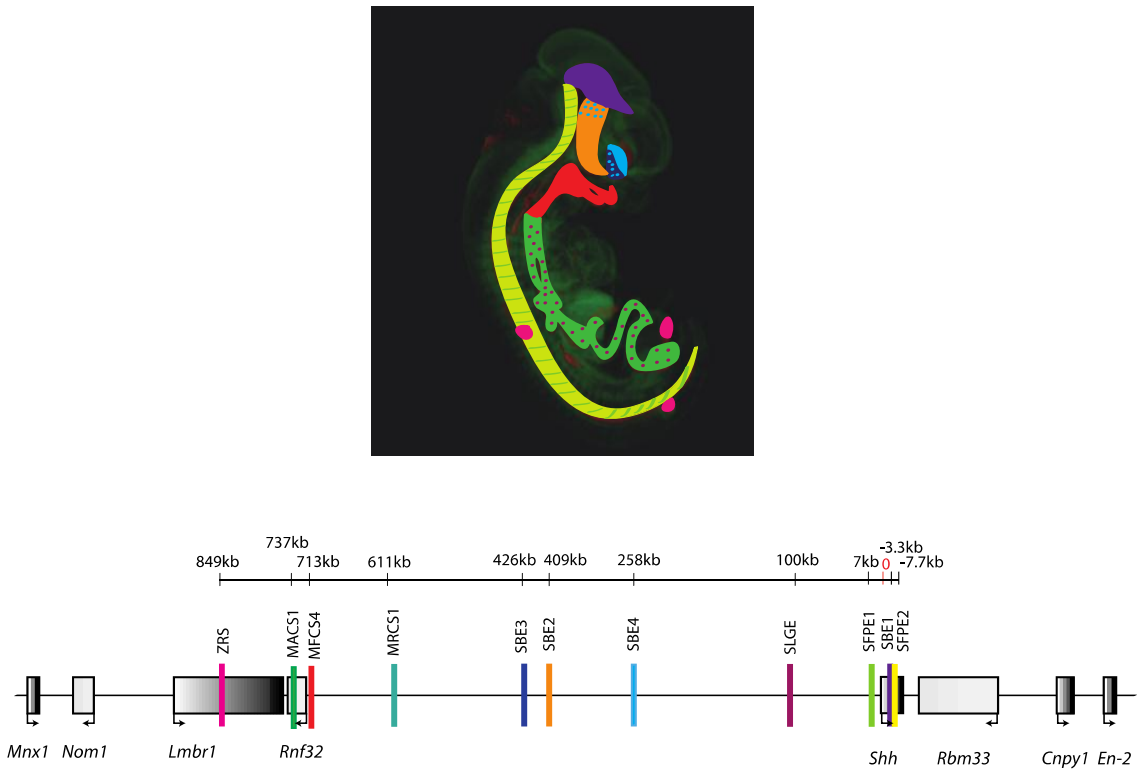


Figure 2.3: Regulation of Shh expression in development

Schematic illustration of the multiple sites of *Shh* expression in the E11.5 embryo (top); the colours used match the relevant enhancers (coloured bars) which are shown in their genomic context alongside other genes within the region (Bottom).

2.8 The *Shh* gene desert

The region between the *ZRS* and *Shh* comprises a large gene desert within which a number of other potential *cis*-regulators of *Shh* reside or flank (Jeong et al. ; Jeong et al. 2006; Sagai et al. 2009) (Figure 2.3). By utilising a transgenic reporter assay in mice two intronic enhancers, *SBE1* and *SFPE*2, and a further enhancer 7 kb upstream of *Shh*, *SFPE1*, were identified (Epstein, McMahon, and Joyner 1999). *SFPE1* and *SFPE2* were found to function redundantly to direct LacZ expression within the floor plate of the spinal cord and hindbrain (Jeong et al. 2006).

SBE1 is capable of driving expression within the ventral midbrain and caudal region of the telencephalon (Epstein, McMahon, and Joyner 1999). A mouse line

has been generated to carry a deletion of the 525 bp intronic sequence mediating SBE1. Embryos lacking *SBE1* were found to initiate Shh transcription but are unable to maintain it within the ventral midline of the rostral midbrain and caudal diencephalon from E10.0 onwards, however reporter gene expression within the forebrain tissue *zona limitans intrathalamica* (*Zli*) of Shh $^{\Delta SBE1/\Delta SBE1}$ was unaffected Together this data suggests SBE1 is important for maintenance of expression in the ventral midbrain (but not the *Zli*) at E11.5 therefore suggesting that *SBE1* could be functioning redundantly with an unidentified enhancer to control *Shh* transcription at earlier time points (Jeong et al. 2011).

A comparative sequence analysis approach was also utilised to examine the region of DNA upstream of *Shh*. Three enhancers were identified 300-450 kb upstream of *Shh*, *SBE2*, *SBE3* and *SBE4* which were found to direct *LacZ* expression within the rostral diencephalon, telencephalon and ventral diencephalon, respectively (Jeong et al. 2006).

A point mutation within *SBE2* which results in a cytosine to thymine base change has been identified within an individual with HPE3 (Jeong et al. 2006). This individual exhibited microcephaly, midfacial hypoplasia, cleft lip and palate, diabetes insipidus and moderate fusion of the hypothalamus and basal ganglia. This base change was shown to result in an inability to drive expression within structures from the optical vesicles to the rostral diencephalon, a phenomenon which was found to be similar to embryos carrying a version of SBE2 which had 10 bp deleted within the region of the base pair change (Jeong et al. 2006). This 10 bp region was therefore hypothesised to be an important binding site for a transcriptional regulator which was later identified as the homeodomain protein Six3. Six3 is therefore believed to function as a context dependent activator or repressor of target gene expression in the developing eye and forebrain, and has also been identified as a mutated locus within some HPE patients (Jeong et al. 2008).

A further cluster of cis-regulatory elements have also been found located 600-900 kb upstream of *Shh*, these elements have been found to reside within and close to

the *Rnf32* gene (see Figure 2.3) (Sagai et al. 2009). Mammal amphibian conserved sequence (*MACS1*) which is located within intron 8 of the *Rnf32* gene was found to drive expression within the respiratory and gut endoderm and caudal regions and also the urogenital tract. Whereas Mammal Fish conserved sequence (*MFCS4*) located just downstream of *Rnf32* was found to direct reporter expression within the oral epithelium, epiglottis, soft palate and arytenoids swelling. Both *MACS1* and *MRCS1* were found to drive expression within an overlapping border in the pharyngeal structures. Finally, Mammal Reptile Conserved sequence 1 (*MRCS1*) was found to act as an enhancer for promoting expression within derivatives of the oral epithelium including the epithelia of the hair and whisker buds, dental placode, rugae of the hard palate and fungiform papillae of the anterior tongue. These elements were therefore shown to drive reporter gene expression in a co-linear pattern along the anterior-posterior body axis within the epithelial linings of the oral cavity and gut (Sagai et al. 2009).

Deletion of *MFCS4* has been undertaken, with homozygote offspring found to be viable at birth, however dying after a few days. These offspring were found to have bloated bellies and have stomachs devoid of any milk, suggesting an inability to swallow. The soft palate within the neonates was found to be truncated, the epiglottis was found to be lost or reduced in size and hypotrophy of the aryteroid and tongue deformation was also exhibited; thus, suggesting that this enhancer is essential for morphogenesis of the pharyngeal structures necessary for respiration and swallowing. Shh expression was also found to be lost within associated regions within mutants as compared to wild-type controls (Sagai et al. 2009).

Recently, an additional enhancer has been located approximately 100kb from *Shh* (See Figure 2.3). This enhancer designated the *SLGE* (Shh lung gut enhancer) has been shown to drive expression similarly to the *MACS1* enhancer. Expression is found within the intestine, oesophagus and whole stomach as well as the alveoli of the lungs. However, unlike *MACS1* it is not capable of driving expression within the laryngotracheal tract. This enhancer is believed to contribute to SHH gut expression

at later stages of development (E12.5) when *MACS1* is not thought to function (Tsukiji et al. 2013) (Figure 2.3).

2.9 Conservation of *Shh* enhancers

The *Zrs* is conserved in a number of vertebrate species including fish and Chondrichthyes. Like mammals, the genomic location of the *Zrs* within the *Lmbr1* gene is also found to be conserved within fish and sharks. This conservation over nearly 400 myr suggests that not only is the sequence of the *Zrs* essential for correct function but also its location in the genomic landscape (Lang et al.).

The oral epithelium-specific enhancer *MRCS1* is conserved in birds and reptiles at the almost same level as between mammals; whereas, its homolog has not been identified in amphibians and teleost fishes. *MFCS4* is conserved between mammals and teleost fishes, although with lower sequence similarity, a homolog has also been identified within Fugu (Woolfe et al. 2005). The lung-gut epithelium-specific enhancer *MACS1* is conserved in amphibians, but not in teleost fishes (Sagai et al. 2009).

SBE2 sequences have been identified in a number of species, with those sequences identified in humans, chicken and amphibians capable of driving expression within the correct tissues in mice. In contrast the Fugu *SBE2* sequence was not capable of this. While *SBE3* sequences were only found in closely related organisms, including chimp and rat. *SBE4* sequences have also been identified in a number of species; however the degree of conservation is reduced in comparison with *SBE2*, with conserved reporter activity from the chick but not the frog or Fugu (Jeong et al. 2006).

It is likely that the *Shh* enhancers have evolved as species diverged with their appearance relative with the presence of the tissues they are expressed in. For example, the *MACS1* enhancer is not present within fish where lungs are also not present. The forebrain enhancers appear to be conserved only functionally within

mammals and birds suggesting they may have arisen relatively recently in accord with more complex brain function. Alternatively, it could be that more diverged species have evolved different enhancers for driving *Shh* expression which have so far not been identified. *SLGE* appears to have very low evolutionary conservation between species, with it, in general, only being present within rodent species, with even rabbit sequences unable to drive expression within the mouse (Tsukiji, Amano, and Shiroishi 2013). It is speculated that this is as a result of the *SLGE* enhancer being of redundant function with *MACS1* and therefore under less evolutionary constraint (Tsukiji, Amano, and Shiroishi 2013).

2.10 Mutations affecting *Shh* regulation

Homozygous deletions of the *Zrs* result in a complete loss of Shh expression in the limb resulting in ulnar aplasia and oligodactyly (Sagai et al. 2005). Heterozygous point mutations within the *ZRS* in both humans and mice have been shown to result in mis-expression of SHH leading to PPD as previously discussed (Lettice and Hill 2005). So far, fifteen different *ZRS* single-point mutations have been identified within humans which are associated with limb abnormalities with the most prevalent being gain of function mutations. Point mutations have also been identified in other species including mice, cats and chickens, where again they are associated with extra toes (Anderson et al. 2012). Point-mutations within the *Zrs* have been shown to result in ectopic anterior Shh expression, which is expected to effectively generate an additional ZPA. This affects the Gli3A:Gli3R ratio leading to the respecification of extra-digits or the transformation of thumb to finger (Anderson et al. 2012). The autosomal dominant disorder Werner mesomelic syndrome has also been shown to be caused by several point mutations within the *ZRS*. This condition presents with a severe phenotype consisting of preaxial polydactyly of the hands and feet, as well as dwarfism as a result of tibial hypoplasia (Wieczorek et al. 2010).

Duplications which include the *ZRS* have been described as causing Triphalangeal thumb polysyndactyly syndrome (TPTPS). Families with Haas Type

polysyndactyly (complete syndactyly with polydactyly), Laurin-Sandrow syndrome (mirror image duplication of all digits) and Syndactyly type IV (SD4) (complete syndactyly affecting fingers of both hands) have also have also been found to occur as a result of duplications of the *ZRS* (Sun et al. 2008). A correlation between smaller duplications and more severe phenotype has been identified; whereby small micro duplications within the *ZRS* region (<80 kb) have been shown to be the underlying genetic cause of Laurin-Sandrow syndrome whereas larger duplications result in less severe phenotypes (Lohan et al. 2014). It is suggested that it is possible that the different sizes of duplications affect the efficiency with which the *ZRS* region interacts with *Shh*. With the sizes of the aberrations affecting the locus dynamics and thereby influencing *Shh* expression negatively, potentially resulting from a similar pathogenic mechanism as *ZRS* point mutations (Lohan et al. 2014).

Chromosomal inversions with breakpoints between *Shh* and the *ZRS* have also been identified both in mice and in humans. Within *Dsh* (short digits) mice, *Shh* has been shown to be ectopically activated in the cartilage of early digit primordia (Niedermaier et al. 2005). It has been suggested that the misexpression of Shh in these mice is the result of the removal of a repressor that enabled additional expression to occur in the early developing digits. Within humans, a TPTPS patient was identified who had an intrachromosomal inversion within the *SHH* region. This led to a phenotype whereby fusions of all fingers and toes along the entire length of each digit had occurred. 'Enhancer adoption' was the term used to describe this phenomenon, as one breakpoint was found to have occurred upstream of the *SHH* gene meaning that it ended up under the influence of a different enhancer at the other end of the breakpoint, freeing it from the influence of the *ZRS* and other regulators (Lettice et al. 2011). Chromosomal rearrangements within the gene desert flanking *SHH* have also been identified. These have been identified as disturbing the *SHH* CNS enhancers resulting in a variable HPE phenotype, similar to that exhibited within *SHH* loss of function mutants (Belloni et al. 1996).

2.11 Neighbouring genes to *Shh*

A number of other potential regulatory targets are present within the *Shh*/*Zrs* genomic region (See Figure 2.3). The *Rnf32* gene lies between the *Zrs* and *Shh* gene, originally suggested to be testes specific is actually found to be ubiquitously expressed during embryo development (van Baren et al. 2002). As previously mentioned (Section 2.8), within its introns enhancers of *Shh* gut and pharynx expression are found (Sagai et al. 2009). The actual function of *Rnf32* however remains unclear and no mouse mutants have been reported to date.

The *Zrs* is located within intron 5 of *Lmbr1*(Figure 2.3), a gene which is expressed ubiquitously within the developing embryo but is not expected to play a role in limb development (Hill 2007). *Lmbr1* deletion mice are found to be viable with a low incidence of limb malformations (Clark et al. 2001).

Upstream of *Lmbr1* (and *Shh*) lie two genes *Nom1* and *Mnx1* (Figure 2.3), very little is known with regards to the former, *Mnx1* however has been shown to play pivotal roles in both pancreatic development (Thaler et al. 1999) and the establishment of motor neuron identity (Thaler et al. 1999). Smaller sized mice which have a 'curled appearance' have been found to result when homozygous *Mnx1* deletions are generated. These mice die at birth as a result of respiratory failure due to improper motor neuron specification (Thaler et al. 1999).

Downstream of *Shh* lie three genes, *Rbm33*, *Cnpy1* and *En2*, *Rbm33* is postulated to be an RNA binding protein. *Cnpy1* is a homolog of a zebrafish gene *Canopy1*. *Canopy1* is expressed in the midbrain-hindbrain boundary in zebrafish, binds Fgfr1, and plays a role in Fgf signalling (Hirate and Okamoto 2006). *En2* is a homeobox containing gene which is implicated in CNS development in humans and mice. Both *Cnpy1* and *En2* have been shown to be expressed within the mid-hindbrain boundary region within the mouse with *En2* playing a role in restricting the fate of progenitor cells to a midbrain or hindbrain lineage (Davis et al. 1988; Paek et

al. 2012). Whether *Cnpy1* also plays a functional role within this region or whether some kind of bystander effect is at play is unknown.

In early stages of limb development, neither *Lmbr1* nor *Rnf32* are up regulated in the ZPA, implying that the *Zrs* is specifically acting on *Shh* (Amano et al. 2009). However, at later stages, *Lmbr1* expression is detected, at least in chicken, in the posterior limb together with *Shh* (Maas and Fallon 2004). This could suggest a relaxation of any mechanisms that ensure a preferential interaction between the Zrs and the *Shh* gene during early limb development. In mouse limbs despite the number of more proximal targets of *Rnf32, Lmbr1, Nom1* and *Mnx1,* the *Zrs* has been found to act specifically on *Shh,* with deletion of the enhancer having no effect on the expression of the other genes (Amano et al. 2009).

2.12 Long range Regulation

The utilization of a number of discrete regulatory regions distributed over large genomic distances is becoming increasingly recognised as a method of organizing temporal and spatial transcriptional gene control. Whole genome comparison and large screens for sequences with regulatory potential has identified enrichment for conserved non-coding elements and enhancers in ‘gene deserts’ located large distances from their nearest gene (Woolfe et al. 2005; Pennacchio et al. 2006).

2.13 Conservation of Gene deserts

The vertebrate genome has been shown to have an uneven distribution of genes, with some regions of the genome found to be gene dense whereas others are found to reflect ‘gene deserts’. Some gene deserts such as that associated with *Shh* contain regulatory sequences that act at large distances to control the expression of neighbouring genes. In contrast other gene deserts are non-essential and can be

deleted without any major phenotypic effects (Nobrega et al. 2004). Comparative sequence analysis of human gene deserts and orthologous regions within the chick genome has divided these two types of gene desert into stable and variable categories (Ovcharenko et al. 2005). Stable gene deserts display high levels of sequence similarity in humans and chickens, whereas variable deserts appear to be specific to mammalian lineages. Stable deserts were found to display lower repeat density and an amount of human/mouse sequence conservation comparable to that of gene rich regions within the human genome (Ovcharenko et al. 2005). They were also found to be depleted of breakpoints disrupting conserved synteny suggesting that a considerable degree of purifying selection has been acting on stable gene deserts. The majority of genes flanking stable gene deserts were found to be those involved in processes such as regulation of transcription, skeletal and muscle development, DNA binding and regulation of metabolism (Ovcharenko et al. 2005).

The function of variable gene deserts is more ambiguous however, it is suggested that these gene deserts could possibly represent recently evolved regions that have not been fixed, or they may lack important function and represent genomic 'junkyards' (Ovcharenko et al. 2005).

2.14 Long range regulation in Development

Like *Shh*, many developmentally important genes are involved in multiple processes during embryo development. The role of long-range regulatory elements appears to be of great importance in regulating developmental genes and several developmental genes have been found to have highly-conserved *cis*-regulatory elements and complex regulatory genomic landscapes associated with them, the importance of which is reflected in their role in common human disorders. In general the regulatory architecture of a single gene may consist of multiple elements that reside within introns and extend for large distances at either end occupying positions in the gene deserts and even in neighbouring genes.

2.14.1 Sox9

Similar to *Shh* other genomic loci have been shown to be composed of regulatory domains comprising multiple enhancer elements responsible for driving expression within separate tissues. SOX9 is an evolutionary conserved transcription factor that is expressed in a variety of tissues. Mutations in the human gene have been shown to result in cause the skeletal malformation syndrome campomelic dysplasia (CD) with associated XY sex reversal (Foster et al. 1994; Wagner et al. 1994). *Sox9* has been shown to be a regulator of chondrogenesis and testogenesis in the mouse (Bi et al. 1999; Akiyama et al. 2002; Chaboissier et al. 2004). CD patients also show defects in other organ systems where *SOX9* is expressed (Ng et al. 1997) such as the inner ear, brain, pancreas and heart (Houston et al. 1983; Mansour et al. 1995; Mansour et al. 2002). Among the symptoms seen in CD are deafness, Pierre Robin sequence (micrognathia, cleft palate and glossoptosis) and scoliosis. Translocation breakpoint analysis within CD patients suggested the presence of multiple *cis*-regulatory elements residing within a 600 kb interval extending 350 kb 5′ and 250 kb 3′ to *Sox9* (Wunderle et al. 1998). Five highly conserved sequence elements were later identified by sequence analysis which lie up to 290 kb 5′ to the gene, with three more, E6–E8 lying up to 450 kb 3′ to human *Sox9* (Bagheri-Fam et al. 2001). These elements were tested within mice with the *lacZ* expression patterns shown to be mediated by the *Sox9* promoter (neural tube and hindbrain), enhancer E1 (node, notochord, gut, bronchial epithelium and pancreas), enhancer E3 (Cranial neural crest cells and inner ear) and enhancer E7 (forebrain and midbrain) shown to recapitulate the many aspects of the complex endogenous *Sox9* expression pattern. All four *cis*-regulatory sequences initiated *lacZ* expression at stages very similar to endogenous *Sox9*. However, while fragments E5–E3 and enhancer E1 were able to recapitulate *Sox9* expression also at later stages, both neural enhancers could reproduce only the early aspect of *Sox9* expression. Elements E2, E4 and E6 did not show independent enhancer activity suggesting these elements could be general enhancers that up-regulate an already specified expression (Bagheri-Fam et al. 2006). More recent evidence suggests a large genomic landscape surrounds *Sox9,*

translocation breakpoints in families with Pierre Robin have been identified both 1.5 Mb telomeric and centromeric to *Sox9*. These have been investigated using a *LacZ* reporter assay which identified the presence of a potential craniofacial enhancer; HCNE-F2 located within a micro deletion region centromeric to the gene (Benko et al. 2009).

2.14.2 Pax6

Pax6 is a transcription factor which plays an important function during development and maintenance of the eye, pancreas and the central nervous system. Haploinsufficiency for PAX6 function has been shown to result in eye defects both the mouse (*Small eye*) and human (aniridia) (Jordan et al. 1992; Glaser et al. 1994). The *Pax6* gene is surrounded by a large number of evolutionary conserved regions (ECR) and a number of cis-regulatory elements have been identified both upstream and downstream of the gene, including within introns of the ubiquitously expressed *Elp4* gene. These regulatory elements have been shown to carry out independent and overlapping functions to drive Pax6 expression (Kleinjan et al. 2006).

Recently, a number of further elements were identified using comparative genomics analysis of the elephant shark genome, including eagBNe6 which was responsible for driving hindbrain and spinal cord activity. The enhancer agCNE1-11 was shown to drive novel olfactory system, pineal gland and nervous system expression. Several of the agCNEs were shown to drive overlapping expression patterns with each other and with previously characterised *Pax6* enhancers including two trigeminal nerve *cis*-elements that both drive highly similar expression patterns, suggesting that in their case the enhancers simply add to the robustness of the regulatory architecture, similar to the role fulfilled by shadow enhancers in *Drosophila*. The full number of *Pax6* enhancer elements are believed to function as one or more archipelagos (Bhatia et al. 2014).

2.14.3 Gremlin

Like *Ssq* which played a role in identifying the *Zrs*, the recessive mouse mutation, limb deformity (*ld*) which results in disruption of the distal limb skeleton

has also played a role in identifying a *cis*-regulatory landscape involved in limb development. The *Ld* phenotype results in synostosis of the zeugopod in combination with oligodactyly and syndactyly of the metacarpal bones and digits. Five separate *ld* alleles were identified, with the *Ld* locus found to consist of a large ~450 kb chromosomal landscape which encompasses two functionally unrelated genes *Formin1* (*Fmn1*) and *Gremlin 1* (*Grem1*). The former is a gene which plays a role in remodelling of the actin cytoskeleton while the latter has been shown to inhibit BMP-2, -4, and -7 serving to regulate the expression of FGFs 4 and 8 and SHH and thus necessary for proper limb development (Gazzerro et al. 2005). Three out of five of the *ld* alleles were found to disrupt the C-terminus of *Fmn1*. However, it was not possible to recapitulate the *ld* phenotype via targeted inactivation of the *Fmn1* gene. Instead deletions of exons 10-24 of *Fmn1* or loss of function mutations of *Grem1* were found to reproduce the associated limb phenotype. It was therefore postulated that a cis-regulatory element associated with *Grem1* regulation was present within the C-terminal of *Fmn1*. The position of this element was later mapped to exons 19-23 of *Fmn1*, located at a distance of 41 kb from *Grem1* (Zuniga et al. 2004; Zeller and Zuniga 2007).

2.14.4 The HoxD cluster

Like *Shh*, the Hox genes, which consist of thirty-nine genes organised into four clusters, play a key role in limb development. The HoxD cluster is necessary for limb development of both the proximal and distal limb segments (Zakany and Duboule 2007). Two phases of *HoxD* transcription take place; an initial phase during early limb budding which involves the activation of the *HoxD* genes at the 3' end of the cluster. This is later followed by a second phase of transcription consisting of expression of the *Hoxd10-Hoxd13* genes in the most distal part of the limb which is important in the development of the digits (Spitz et al. 2003). Similarly to *Shh*, the *HoxD* cluster is also part of a complex regulatory landscape. Two distinct enhancer systems have been identified located either side of the HoxD cluster, referred to as conserved sequences B (CsB) and CsC (Spitz et al., 2003 and Gonzalez et al., 2007), these elements are capable of driving reporter gene expression in digits. CsB is part

of a Global Control Region (GCR) conserved in all vertebrates and containing various enhancers. It is located 180 kb upstream *Hoxd13*, in a 600kb large gene desert extending from *Lunapark (Lnp)* until *Atp5g3*. CsC is part of the Prox enhancer, located between *Lnp* and *Evx2 (Gonzalez et al. 2007)*, which are both coexpressed with *Hoxd* genes in digits as a bystander effect (Spitz, Gonzalez, and Duboule 2003). The combined effect of CsB and CsC is proposed to be required for proper activation of *Hoxd* genes in digits (Gonzalez, Duboule, and Spitz 2007).

2.14.5 Fgf8

Fgf8 encodes a secreted signalling molecule expressed within a number of signalling centres within the developing embryo. Mice lacking Fgf8 die before E8.5 as a result of a gastrulation defect (Meyers et al. 1998; Sun et al. 1999). Experiments on later embryos has identified that Fgf8 also plays a role in the development of the mid-hindbrain region, the limbs, the brachial arches, heart, kidney and the brain (Lewandoski et al. 2000; Moon and Capecchi 2000; Sun et al. 2002; Chi et al. 2003; Garel et al. 2003). The *Fgf8* gene sits within a relatively conserved genomic region which also contains the conserved gene *Fbxw4* (Komisarczuk et al. 2009). A number of regulatory elements were identified in that region; firstly, two 3' to *Fgf8*; *CR3* and *CR4* which drive expression within the AER, mandibular and maxillary arches, somites and the isthmus (Beermann et al. 2006). In addition a number of elements have also been identified within the exons of *Fbxw4* and beyond at a distance of up to 90 kb away from FGF8 itself (Komisarczuk, Kawakami, and Becker 2009).

2.15 Genomic Landscapes

2.15.1 What directs enhancers towards their correct promoter?

The distance at which some developmental enhancers sit and the potential complexity of additional genes within a region raises the question of how enhancers

act on a specific target. Some enhancers have broad specificities acting on more than one target within a region such as reported for HoxD cluster whereby the limb bud expression of the genes *Hoxd10-13*, and the neighbouring Lunapark (*Lnp*) and *Evx2* genes, are controlled by the same group of regulatory elements (Spitz, Gonzalez, and Duboule 2003). This phenomenon is often termed as a 'bystander effect' and is found at a number of loci including *Lmx1b, Dlx5/Dlx6* and *hGh* and *Cd79b* (Crackower et al. 1996; Holmes et al. 2003; Cajiao et al. 2004).

The organisation of DNA is important for expression of genes. Within the nucleus genomic DNA is found packaged around histones and other larger chromatin structures. In general, compact DNA (heterochromatin) is believed to be associated with repression of gene expression by preventing transcription factors and RNA polymerases from accessing enhancer and promoter sequences. In comparison, chromatin found around active genes is more relaxed which increases the chance of access to promoter regions. Post-translational modifications of histones, through acetylation or methylation by chromatin modifying enzymes have been shown to be associated with an increase or decrease of transcriptional activity (Li et al. 2007). The presence of *cis*-regulatory regions could be important for control of chromatin domains. Remote regulatory elements can function to establish extensive 'activated chromatin domains'. This has been demonstrated at both the human and mouse *β-globin* locus whereby a locus control region (LCR) and intergenic regions are transcribed by RNA Polymerase II in a subset of erythroid cells *in vivo* which is required to generate and maintain an open chromatin domain (Gribnau et al. 2000).

Often genes located within close proximity at loci are found to be co-expressed. For example; the *Lnp*/E*vx2* or *Formin* genes which have no important function in limb development are strongly expressed within the developing limb bud. This expression however results from their position within either the *HoxD* or the *Gremlin* regulatory landscapes (Trumpp et al. 1992; Herault and Duboule 1996; Spitz, Gonzalez, and Duboule 2003; Zuniga et al. 2004). It was proposed that transcription of these additional genes occurs simply because of their close proximity to a strongly transcribed region, thus appearing as a bystander effect. Such bystander regulation

may either involve a shared enhancer, like the GCR at the *HoxD*, which may contact promoters surrounding strongly transcribed genes (Spitz et al. 2005). Alternately, this could occur as a result of the spreading of transcriptionally permissive chromatin structures around active genes. This is seen at the *hGH/CD79B* locus, where both genes are transcriptionally active in the pituitary, yet only hGH is found at the protein level, while CD79B is translated in B cells, where it is important for signal transduction from the B-cell receptor (Cajiao et al. 2004). This is proposed to be a non-functional consequence of the genes' localization within the highly acetylated chromatin domain defined by the enhancer of the growth-hormone genes (Cajiao et al. 2004). It has also been suggested that the presence of neighbouring promoters may be important to fine-tune the expression level of the functionally active target gene, not by acting in trans but rather by interactions in cis, for instance through the titration of the regulatory input. The deletion or addition of genes within the *HoxD* cluster has been shown to lead to reallocation of the GCR activity, therefore preventing the action of the GCR on *Lnp* and *Evx2* might also redistribute the regulation on the remaining HoxD genes in a way deleterious for limb development (Spitz and Duboule 2008).

Remote regulatory elements can be brought into close juxtaposition with target promoters via long-range 'looping' or 'tracking' mechanisms. Chromosomal regions containing transcriptionally active genes tend to loop out of the condensed chromosomal territory, coming close to the periphery of the nucleus in 'transcription factories'. Genes that are co-expressed but separated by several megabases have been shown to co-localise in discrete nuclear compartments (Osborne et al. 2004). By measuring physical distance using 3D-FISH within cells from the ZPA, *Shh* and the *Zrs* were shown to have a much reduced intranuclear distance as opposed to a control region within the desert, suggesting that they interact via a looping mechanism (Amano et al. 2009). This occurs as a two-step process whereby firstly a chromosomal looping brings the Z*rs* close to the *Shh* promoter and interacts with it, *Shh* then moves out of its own chromosomal territory which is expected to relate to the genes activation. Deletion of the *Zrs* from its genomic region results in a dramatic reduction of Shh expression within the limb bud, however it was proposed that co-

localisation of both the *Zrs* and *Shh* was still found to occur within ZPA nuclei. This however may not be the case as a more recent study suggests a direct relationship between *Shh* gene expression in the limb bud and chromosomal conformation, at E11.5 a 3´del of the *Zrs* results in forelimb expression inactivity but low level expression activity in the hind limb. A direct correlation between chromosome conformation measured by FISH and enhancer activity was found. Amano's previous work was performed within earlier staged embryos (E10.5). Therefore, a complex picture is suggested in which the *Zrs* at the earliest stages of limb bud development is driven to interact with the *Shh* gene by as yet, unidentified element(s) but this task is subsequently acquired by the *Zrs* itself by E11.5. The 3´ half of the *Zrs* operates by boosting activity within this mechanism that mediates these enhancer/promoter interactions (Lettice et al 2014).

Shh and the *Zrs* are shown to physically interact even in cells in which *Shh* is not actively being expressed such as those of the anterior of the limb bud. It is therefore proposed that the locus is possibly being held in three states; silent in which they do not contact, poised in which there is some contact and then active where the gene is transcribed. As the anterior of the limb bud does have the ability to express Shh, the poised state of the chromatin within these cells is believed to serve as a sign of the competence of the cells (Amano et al. 2009). Downstream promoter elements, *TATA* boxes and other tethering elements are implicated in directing promoter-enhancer interactions, suggesting that the core promoter mediates enhancer interactions and could determine specificity of the enhancer (Butler and Kadonaga 2001; Calhoun et al. 2002). It could be that these elements play this role by facilitating or stabilising chromosomal loops.

Some genomic landscapes are more complicated than others. The *HoxD* landscape in particular has been shown to be complex with multiple enhancer elements spread over 800 kb of gene desert. This region has been designated to function as an archipelago whereby multiple element islands are brought into the vicinity of *Hoxd13* via chromatin looping. In the digits these islands contact each other leading to multiple simultaneous interactions between distinct elements. In

such an instance, it is hypothesised that transient interactions would also occur between islands, rather than only between enhancers and promoters (Figure 2.4). In tissues where the gene is not expressed (such as the brain) a less elaborate structure may occur in a default or poised position. Whereas, in the limb it is speculated that additional specific factors may initiate transcription such as chromatin modifying factors and RNAPII to active chromatin loops (Montavon et al. 2011).

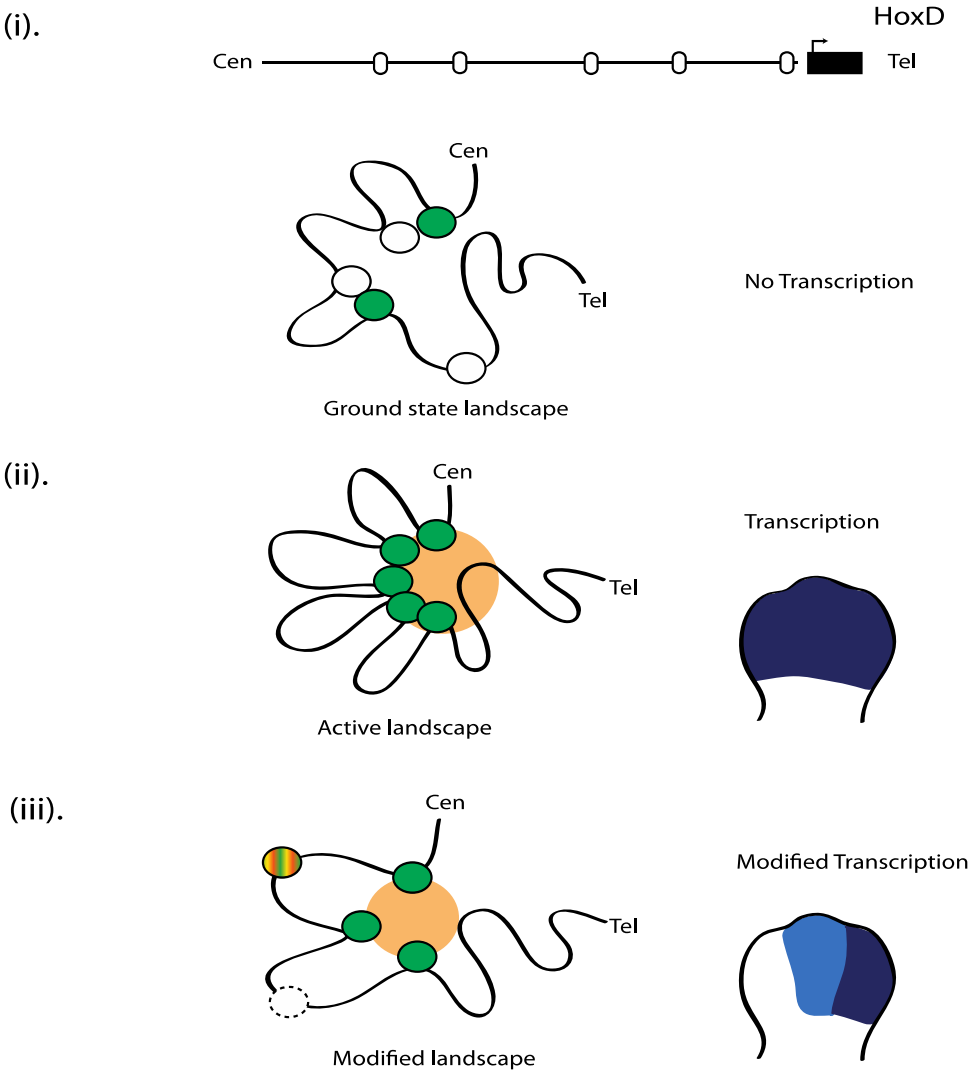


Figure 2.4:HoxD Regulatory Landscape

Schematic representation of transcriptional output of HoxD locus (i) the regulatory landscape covering approximately 800 kb upstream of the HoxD cluster. Multiple regulatory islands (ovals) are required for HoxD activation in limbs, either by acting as enhancers or anchoring points. When HoxD genes are not expressed, the landscape is in a ground state with only some regulatory islands contacting but insufficient to activate transcription. (ii) Additional contacts are made in expressing cells, islands become active (green) as a result of histone marks and recruitment of RNAPolII. HoxD is expressed

throughout the limb. (iii) Modifications to the islands in terms of spacing or order results in a modification to the amount and spatial expression of the gene resulting in pathological modifications.

Experimental approaches have shown that the mere position of a gene with respect to an enhancer can be important for function (Tarchini and Duboule 2006). The importance of the genomic context has been shown at the *Fgf8* locus. A number of enhancers are present within the *Fgf8* region; however these do not all drive *Fgf8* expression. It is suggested that *cis*-interactions between these different modules may reinforce the frequency of their association with the target gene (Marinic et al. 2013). The multiple enhancers in proximity to *Fgf* could lead the region to adopt a defined structural conformation that could favour regulatory interactions between enhancers and specific positions of the locus instead of specific promoter sequences, thus filtering out the activity of non-essential enhancers (Marinic et al. 2013). The *FGF8* locus is therefore proposed to behave as a 'holoenhancer'. Holoenhancers are a complex of regulators that together control a complex developmental expression pattern. It is suggested that the regulatory locus cannot be defined by its basal components with instead the output of the entire region determined by the relative distribution of genes and enhancers along the locus (Marinic et al. 2013) (Figure 2.5).

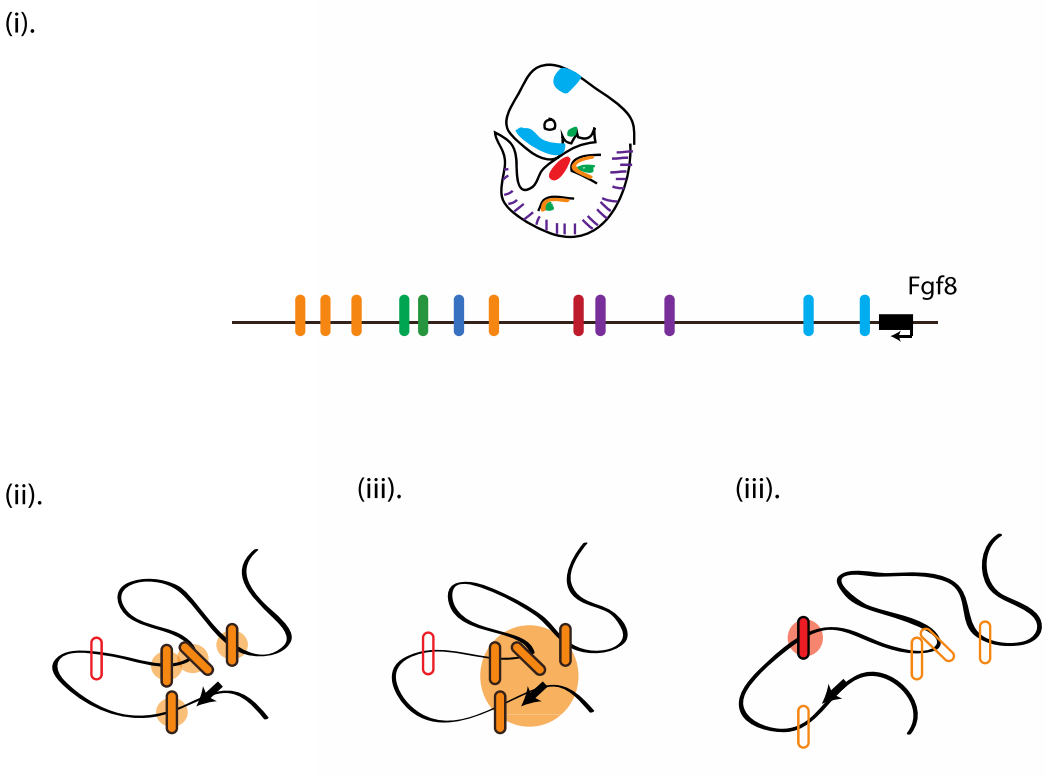


Figure 2.5: FGF8 Regulatory Landscape

(i) The *Fgf8* gene region comprises multiple enhancers (coloured rectangles) which drive expression within the mouse embryos (shown above colour coded).

(ii) In *Fgf8* expressing cells multiple modules with overlapping activities act together either by their own dynamic one-to-one interactions or (iii) by combining their activities to form a holoenhancer (iv) in these scenarios other elements active in other cells types (red) may only have short range activities and therefore minimal influence.

Enhancer status is indicated (filled, active, empty not active) with halo representing their range of action. Transcriptionally active genes represented by black arrow.

2.15.2 What prevents enhancer activity?

Boundary elements, or insulators, have been defined in part by their ability to block enhancer activity when located between an enhancer and a promoter (Geyer 1997). The most well-characterized examples of such elements include *SCS/SCS8*, originally identified in the *Drosophila 87A7* heat shock locus, the *Gypsy*

retrotransposon, also in *Drosophila*, and in vertebrates *58 HS4* from the chicken β-globin LCR (Chung et al. 1993).

Data from Hi-C experiments (a method for identifying higher order chromatin interactions genome wide) suggests the genome of mice and humans is organised in large chromatin interaction domains known as 'topological domains'. These domains are believed to correlate with regions of the genome that constrain the spread of heterochromatin and consist of highly self-interacting regions in which tens of genes and hundreds of enhancers reside (Dixon et al. 2012). It is speculated that these domains could constrain looping interactions between enhancers and promoters and set the boundaries of coordinated gene regulation, and indeed *Shh* and the *Zrs* are found to occupy the same topological domain (Smallwood and Ren 2013). At the edges of these topological domains there are narrow segments in which chromatin interactions appear to end abruptly. These regions are believed to represent 'boundary regions' which are enriched for the CTCF binding sites, housekeeping genes, transfer RNAs and short interspersed element transposons (SINEs) and believed to share the enhancer/promoter blocking feature of classical 'insulators'.

CTCF is a highly conserved DNA binding zinc finger protein that displays nearly 100% homology between humans, mice and chick. It is critically important during a variety of cell processes, shown by the fact that CTCF homozygous mutant mice exhibit early embryonic lethality (Heath et al. 2008). Maternal depletion of Ct has also been shown to disrupt progression to the blastocyst stage (Fedoriw et al. 2004). In adult animals, CTCF is found to be ubiquitously expressed similar to housekeeping genes. CTCF has been implicated in a diverse number of roles involving gene regulation including transcriptional activation and repression as well as genomic imprinting. CTCF is also speculated to play a role in enhancer blocking and/or barrier insulation i.e. blocking communication between adjacent regulatory elements in a position-dependent manner or buffering genes from position effects caused by the spread of heterochromatin. This was initially demonstrated within the *β-globin* locus. At the 5´ end of the *β-globin* locus sits a DNAse1 hypersensitive site (5' HS4) which separates the locus from neighbouring heterochromatin, CTCF is

found to bind to HS4 and has been demonstrated to be responsible for insulator activity (Farrell et al. 2002; Bulger et al. 2003). An accumulating amount of data suggests that CTCF is responsible for mediating long-range chromatin interactions between insulator elements, in a manner similar to insulators in *Drosophila* (Gerasimova and Corces 2001). 3C has been used to demonstrate at the *β-globin* locus that the CTCF sites present physically interact with each other, this causes formation of large chromatin loops just before gene activation encompassing the β-*Globin* regulatory control region (*LCR*). It is speculated that these loops then facilitate subsequent spatial interactions between the *LCR* and its genes (Phillips and Corces 2009). However analysis of CTCF within the *HoxD* locus suggests that despite this paradigmatic situation for CTCF acting as an insulator protein, the conditional loss of CTCF within the forelimbs does not a show evidence for such a function of CTCF, at least in developing limb mesenchyme (Soshnikova et al 2010).

It is unlikely that CTCF sites alone are capable of acting as enhancer blockers. Only 15% of CTCF sites are associated with the boundary regions of topological domains. The rest are found distributed across the genome including in regions interlaced between promoters and enhancers (Dixon et al. 2012). Many models of enhancer landscapes suggest dynamic interactions between different modules during looping processes, the role of CTCF alongside Cohesin could therefore be to stabilise these interactions to allow gene expression within intended tissues. This is also likely to involve proteins involved with histone modifications, chromatin-binding proteins and transcription factors as well as tissue specific proteins.

2.16 The Local Hopping Enhancer Detection System

Identifying the sequence elements and the combinatorial rules that determine enhancer function is necessary to fully understand how enhancers direct the spatial and temporal regulation of gene expression.

A general accepted hypothesis is that functionally important regulatory sequences are under purifying selection. As a result, conserved noncoding sequences (CNSs) were seen as natural candidates for enhancer elements. Early studies used CNSs to detect putative enhancers and test their activity in zebra fish or mouse reporter assays (Woolfe et al. 2005; Pennacchio et al. 2006; Visel, Blow, et al. 2009). Both BACs and then YACs have also been utilised for the identification and analysis of regulatory sequences found at long distances (Nielsen et al. 1997; Kaufman et al. 1999). While BACs provide a limited 300 kb (in the case of regulatory loci) cloning capacity YACs are capable of the stable maintenance of genomic fragments of DNA larger than 1 Mb (Burke et al. 1987); however, use of either for transgenesis requires laborious specialised techniques and may still not be large enough to address some genomic loci. Although these conservation-based approaches achieve some success, limitations also exist. The function and spatio-temporal specificity of CNSs cannot be determined by conservation alone and, therefore, requires additional experimentation. In addition, several studies have shown that noncoding sequences that apparently lack conservation (as assessed by sequence alignment) may still contain functional regulatory elements (Fisher et al. 2006; Birney et al. 2007; McGaughey et al. 2008).

The mouse represents an excellent model for analysing functional genomics either by genome-wide random mutagenesis by a chemical or insertion mutagen, or pinpointing a single locus by gene targeting techniques. However, again until recently only limited tools have been available for the survey of large regulatory regions. These either involved the sequential targeting of *LoxP* sites into the same chromosome and scanning via a ‘walk and delete’ manner in comparison is very labour-intensive. The alternative strategy retrovirus mediated random integration which is not suitable for pure insertional mutagenesis with no accompanying deletions being formed, such as single-gene knockouts or enhancer traps in a native genomic context. Recently, however several systems have been devised using transposons which can be used to explore regulatory genomes on a large scale (Kokubu et al. 2009; Ruf et al. 2011).

The local Hopping Enhancer Detection (LHED) system is a chromosomal engineering system based on the sleeping beauty transposon which has previously been used to examine the mouse *Pax1* locus. Within the *Pax1* locus it has been employed to show that transcriptional insulators are present between the regulatory gene loci of *Pax1* and its neighbouring gene *Nkx2-2* (Kokubu et al. 2009).

The Sleeping Beauty transposon is a binary system comprised of a transposon vector flanked by inverted repeats/direct repeats which can be mobilised via the addition of a sleeping beauty transposase in *trans*. To generate the LHED system firstly an enhancer detector system was cloned into the transposon DNA. This comprised an *Hsp68* minimal promoter fused to the *LacZ* reporter gene. The transposon-containing the reporter gene was then inserted between a PGK promoter and a puromycin resistance gene so that upon excision of the transposon reactivation of puromycin resistance occurred. Three L*oxP* sites were also inserted within the vector, one within the transposon and two outside. This was to allow chromosomal engineering (deletions or inversions) between a fixed insertion site and the transposition site, once the transposon had been induced to hop. The complete LHED vector therefore combines standard knock in technology, transposon based enhancer detection and *Cre/LoxP*-mediated chromosomal engineering to allow monitoring of enhancer activities along a targeted genome region and generation of a nested series of deletion mutations to examine loss-of-function effects within a genomic neighbourhood (Kokubu et al 2009).

The sleeping beauty transposon has a strong tendency to favour 'local hopping' and reinsert at closely linked loci during transposition. The system also benefits in its use of targeting into ES cells which provides a practical advantage in that it allows clone archiving. Both of these features of the system mean that it is possible to saturate a relatively large genomic distance with enhancer detectors with relative ease (Kokubu et al 2009). Additionally, because the system involves a single copy insertion, it does not suffer from undesirable rearrangements including deletions, inversions and translocations as a result of concatemer formation.

2.17 Aims

2.17.1 Investigating the composition of long-range regulation within the *Shh* locus.

The *Shh* gene is expressed within a large number of developmental tissues during various stages of development where it plays a number of roles. The *Shh* locus presents as a complicated regulatory landscape comprising nearly 1 Mb containing a number of *Shh* enhancers located within the *Shh* gene itself, within a large a gene desert adjacent to *Shh,* as well as several unrelated genes residing within the region (Figure 2.3). These multiple enhancers serve to drive expression of *Shh* within these multiple tissues thus contributing to the highly organised expression pattern of the gene within a given tissue at a given stage. Complicated long-range regulatory domains are a common method of regulating the expression of developmental genes, however, many questions remain regarding exactly how these function with recent data supporting a number of regulatory mechanisms.

The DNA element the *Zrs* which drives *Shh* expression within the limb sits almost 1 Mb away from the gene thus presenting as the furthest known enhancer of *Shh*. The *Zrs* and other *Shh* enhancers operate over large genomic distances and it is speculated that a number of additional enhancers or shadow enhancers reside within this region. This organisation is highly conserved between vertebrates suggesting it could be intrinsic to function (Hill 2007). Recent datas suggest that regulatory loci are often located within topological domains flanked by boundary regions which prevent illegitimate interactions between enhancers and other genes. A proposed domain has been suggested for *Shh,* however it remains unclear exactly where the boundaries of enhancer function exist.

This project will therefore involve investigating the *Sonic hedgehog* region with an aim to investigate several questions about the composition of this regulatory locus;

- Is the distance between the *Zrs* and *Shh* necessary for *Zrs* function?
- Are there elements within the locus that mediate long distance promoter recognition?
- Have *Shh* enhancers evolved activity commensurate with position or with distance?
- Do any previously undiscovered elements/enhancers exist within the region and what is their function?
- What restricts *Zrs* regulation to the *Shh* locus?
- What are the genomic limits of enhancer activity? And how is this activity confined?
- Can systematic deletions within the region affect enhancer activity and thus *Shh* expression?
- Will removal of an overlapping set of enhancers provide further clues to *Shh* function in development?

The Local Hopping Enhancer Detector (LHED) system will be employed, which has previously been tested (Kokubu et al. 2009). I propose to use LHED vectors inserted on either side of the *Zrs* so as transposition can be directed in either direction with a hope of saturating the region. This will allow the *LacZ* reporter gene to be utilised in order to map enhancer activity across the region as well as both upstream of *Lmbr1* and downstream of *Shh*. The orientation of vector design will be taken into account in order to allow *Cre*-mediated deletion of the region between the *Zrs* and *Shh* to allow analysis of the effect of deletions within this region on enhancer activity and also phenotype.

3. Materials and Methods

3.1 Plasmids

3.1.1 DNA BAC

BAC DNA used for amplification was PAC 542-N10 from the RCPI21 library, as described by Osoegawa (Osoegawa et al. 2000).

3.1.2 Plasmids

The pBluescript II KS (-) vector was supplied by Agilent Technologies, and used as described by the manufacturer. The cre plasmid used was pGK Cre bpA supplied by Kurt Fellenberg.

3.1.3 LHED system

The LHED plasmid was supplied as a gift by Chikara Kokubu, the original CMV-SB11 transposase by P.B Hackett and the hypermutated CMV-SB100x by Zsuzsanna Izsvak

3.2 DNA spectrophotometry

DNA concentrations were measured at 260/280 nm absorbances using a Nanodrop 1000 spectrophotometer (Thermoscientific).

3.3 Agarose Gel electrophoresis

Agarose powder was dissolved in 1x Tris-borate-EDTA (TBE) buffer by heating in a microwave and allowed to cool. 1 µg/ml ethidium bromide was added, before pouring the gel into a gel mould and inserting combs. DNA samples were mixed with 6x loading buffer and loaded into the wells; one well was reserved for loading molecular weight markers. An electrical voltage of approximately 100-150 mV was applied and DNA was visualised under UV light.

3.4 Sequencing

Sequencing reactions were performed by the Technical Service Department at the MRC Human Genetics Unit. Sequence analysis was performed using Sequencher 4.8.

3.5 Nucleic acid extraction

3.5.1 DNA extraction from mouse tissues

Tail tips were taken and incubated in 500 µl tail lysis buffer (100 mM Tris-HCl, pH 8.5; 5 mM EDTA; 0.2% SDS; 200 mM NaCl) with 10µl Proteinase K (10 mg/ml) and incubated overnight at 55^{o}C. Samples were centrifuged at 13,000 rpm for 5 min at 4^{o}C. Supernatants were transferred to fresh tubes and 500 µl of isopropanol was added and mixed. DNA was then spooled out and added to 500 µl H20.

Ear clip samples were incubated in 75 µl of Base solution (1.25M NaOH, 10mM EDTA, pH 12) at 95^{o}C for 30 minutes. Samples were then allowed to cool before addition of 75µl of Neutralization solution (2M Tris-HCL pH 5) was added and mixed. 2 µl were then used per PCR reaction.

3.5.2 DNA extraction from ES cells

Cells were expanded to the 24-well stage before being treated with 0.5 ml of lysis buffer (10 mM Tris pH8, 10 mM EDTA, 0.6% SDS, proteinase K) overnight at 37 °C. Following Lysis, 1 ml of ice cold acetone/DMF (70% acetone, 5% dimethyl formamide) was added to lysis mix and vortexed. DNA was spooled out and then washed in ice-cold 70% ethanol for approximately 2 hours-overnight, before being resuspended in 50 µl of H_2O and incubated at 55 °C for several hours to dissolve.

3.5.3 RNA extraction from mouse tissues and cell cultures

RNA was isolated from tissues collected from animals or from cultured mouse embryonic (ES) cells. For cells, media was removed and cells were washed once in sterile PBS. Cells were then resuspended in TRIzol® LS Reagent (Invitrogen, Cat. No. 10296028). For RNA isolation from tissue, fresh tissues were dissected from E11.5 mouse embryos, RNA was then extracted using RNAbee (AMSBio, Cat. No. CS-104B) according to manufacturer's instructions. RNAs were run on agarose gels (1% in TBE) to check for quality and degradation, and yield was calculated with a NanoDropTM 1000 (Thermo Scientific).

3.6 Polymerase Chain Reaction

For all PCR reactions oligonucleotides were rehydrated to 1 µg/µl stock solutions and 20 mM working solutions were set up. Master mixes were made up where multiple reactions were performed. All products were run on agarose gels with 1 Kb molecular DNA size marker (Invitrogen, Cat No. 10787-018 or New England Biolabs, Cat No. #N3232).

3.6.1 Proofreading PCR

PCR reactions were performed as follows 3 µl 1.5 mM $MgCl_2$, 5 µl 10 mM DNTPs, 3 µl of 20 mM primers, 1 µl of 10 ng/µl DNA, 5 µl 10x buffer, 1 µl of KOD hot start polymerase (Merk Biosciences, Cat No. 71086) and water up to 50 µl. This reaction was cycled as follows: 95°C x 3 minutes, followed by 35 cycles of 95°C x 30 seconds; 60°C x 30 seconds, 72°C x 2 minutes, followed by a single extension of 2 minutes at 72°C.

3.6.2 Standard PCR

PCR reactions were performed as follows; 2.5 µl 10x Buffer, 1.5 µl $MgCl_2$, 1 µl 100 mM DNTPs, 4 µl 20 mM primers 0.2 µl of Taq (Invitrogen, Cat No. 10342020) and H_2O up to 25 µl. Reactions were cycled as follows: 94°C x 3 minutes, followed by 30x cycles of 94°C x 20 seconds; 58°C x 30 seconds, 72°C x 1 minute, followed by a single extension of 5 minutes at 72°C.

3.6.3 Long Range PCR

Reagents mixed as follows; 16 µl H_20, 3.5 µl 10 mM DNTPs, 1.5 µl 10 mM primers, 2.5 µl 10xPCR buffer 3, 1 µl Template DNA, 0.3 µl Expand Taq (Roche, Cat No 11681842001). Reactions then cycled as follows; 94°C x 2 minutes, followed by 10x cycles of 94°C x 10 seconds, 58°C x 30 seconds, 68°C x 4 minutes and 25x cycles of 94°C x 15 seconds, 58°C x 30 seconds, 68°C x 4 minutes, followed by a single extension of 68°C x 7 minutes.

3.6.4 Nested PCR

Nested PCR for identification of LHED re-insertion was performed as described by Ruf et al., (2011). A nested asymmetric PCR strategy using Platinum Taq

(Invitrogen, Cat No 10966083) was employed. A first round of PCR amplification was performed on a genomic DNA template from ES cells with a transposon-specific primer pointing outward from either end of the transposon (SB-R1 for the right end, SB-L1 for the left end, 20 μM) and a random primer with a 5-bp 3′ anchor (KmonP-N7-ctcag or KmonP-N7-tcctg, 100μM) (See Table 3.1 for details). A second round of amplification was carried out on 1 μl of the initial PCR reaction using SB-R2 (right) or SB-L2 (left) and KmonP (20 μM). The final PCR product was then sequenced using SB-R3 for the right end or SB-L3 for the left end. These resulting sequences were analysed using Sequencher and aligned to the mouse genome sequence (Build 37, mm9, using the UCSC Blat webpage.

3.7 Single-stranded cDNA synthesis & RT-PCR

RNA was used for single-stranded cDNA synthesis by reverse transcription. The 1st Strand cDNA Synthesis Kit for RT-PCR (AMV) (Roche, Cat. No. 11483188001) was used according to the manufacturer's instructions using oligo-dT primers and with various RNA input quantities. Controls were set up for each sample without enzyme to identify genomic contamination.

3.8 Phenol chloroform/chloroform extraction

1 volume of 25: phenol 24: chloroform 1: isoamyl alcohol was added to each DNA sample and centrifuged (1 minute; 13,000 rpm; 4°C). The DNA layer was removed and added to 1 volume of 24: Chloroform 1: isoamyl, vortexed and centrifuged (1 minute; 13000 rpm; 4 °C). The top layer containing the DNA was then transferred to a clean Eppendorf tube.

3.9 DNA precipitation

1/10 volume of 3M sodium acetate (pH 5.2) and 2.5 volumes of 100% ethanol were added to DNA samples mixed and incubated at -20°C for 1 hour-overnight. Samples were then centrifuged (30 minutes; 13,000rpm; 4°C). The supernatant was removed and the pellet washed in 70% ethanol, centrifugation followed (10 minutes; 13,000rpm, 4°C). The ethanol was removed and the pellet air-dried for 10 minutes before being resuspended in H_2O.

3.10 Restriction Enzyme Digest

DNA digestion was performed as follows; 10 µg of plasmid DNA, 20 µl of buffer (10x), 100 µg of each restriction enzyme required (New England Biolabs or Roche) and H_2O up to 200 µl. This was then incubated at the required incubation temperature for one hour. The reaction was scaled down when required.

3.11 Ligations

Ligations were performed using the Rapid DNA ligation kit (Roche, Cat No. 11 635 379 001), according to manufacturer's instructions.

3.12 Transformation

Transformations were performed using library efficiency DH5α (Invitrogen, Cat No. 18263012) or JM110 (Stratagene, Cat. No. 200239) competent cells, according to manufacturer's instructions.

3.13 Bacterial recombineering

EL350 cells were inoculated into 5 ml of LB broth and incubated at 32°C overnight with shaking. The following day 50 ml of the overnight culture was seeded into 200 ml of LB broth and grown at 32°C with shaking for approximately 4 hours. Cells were then transferred to a 42°C water bath to induce expression of genes required for recombination and incubated for 15 minutes with shaking before being transferred to ice for 30 minutes. Cells were then centrifuged (10 minutes; 3100rpm; 4°C) and resuspended in 1 volume of ice-cold water. Centrifugation was performed again (10 minutes; 3100 rpm; 4 °C) and the pellet resuspended in ½ volume of ice-cold water. This was repeated once more, followed several washes in 10 % glycerol. Finally, the cell pellet was resuspended in 400 µl 10% glycerol and aliquoted into 40 µl aliquots. 40 µl of cells were then transferred to an electroporation cuvette (2mm gap) and 10-50 ng of vector DNA added. Electroporation was performed using a Bio-Rad Gene Pulser under the following conditions 2.5 kV, 25 microF and the pulse control set to 200 Ω. 1 ml of LB broth was then added to each cuvette, which was incubated at

32°C for 1 hour. Cells were then spread on plates containing the appropriate antibiotic.

3.14 Small-Scale Plasmid Purification

Small-scale plasmid purification was performed using the QIAprep Spin Miniprep Kit (Qiagen, Cat. No. 27104), according to the manufacturer's instructions.

3.15 Large-scale Plasmid Purification

Large-scale plasmid purification was performed using the QIAprep Spin Maxi prep Kit (Qiagen, Cat. No. 12162), according to manufacturer's instructions.

3.16 Gel Extraction

DNA extraction from agarose gel was performed using the QIAquick Gel Extraction Kit (Qiagen, Cat. No. 28704), according to manufacturer's instructions.

3.17 ES Cell Culture

3.17.1 Maintenance of ES Cells

Sterile technique was applied during all procedures in the ES cell laboratory. All solutions were pre-warmed to 37 ˚C before use.

The ES cells were grown on gelatinised plates. Plates or flasks were coated with 0.1% gelatine solution (Sigma) and incubated at 37 ˚C for at least 15 minutes before plating. Cells were maintained in 37 ˚C and 5% CO_2 incubator.

3.17.2 E14IVtg2a /G4 ES cell culture

Mouse E14IVtg2a were provided by Austin Smith. Mouse E14IVtg2a /G4 ES cells were cultured in Dulbecco modified Eagle medium (Invitrogen, Cat. No. 11995) supplemented as follows;

Culture media:

Glasgow MEM	500 ml
FCS (Globalpharm)	10%

Glutamine (Gibco)	0.3%
MEM non-essential amino acids (Sigma)	0.1 mM
Sodium Pyruvate (Sigma)	1 mM
β-Mercapthoethanol (Sigma)	0.1 mM

Leukaemia Inhibitory Factor (LIF) was also added in order to keep ES cells in their undifferentiated state; this was produced by Fiona Kilanowski from a plasmid gifted by Austin Smith.

3.17.3 Cell Splitting

Since cells were adherent to the plates, trypsin was used to detach them. First the cells were washed with PBS twice to remove the serum containing medium. They were then incubated in 1 ml of trypsin at 37 °C for several minutes. 9 ml of media was added to the trypsinised cells to stop trypsin activity and the sample was placed in a 15 ml Falcon tube and centrifuged for 5 minutes at 1,000 rpm. The supernatant was removed and the pellet was re-suspended in 1 ml media, dispersed into single cells and re-plated in the desired dilution. Cells were fed on a daily basis and usually split as 1/5 on an every other day basis.

3.17.4 Cryopreservation

2X Freezing media was prepared (20% DMSO (dimethyl sulfoxide), 80% FCS (fetal calf serum) and kept on ice until use. The cells were trypsinized, re-suspended in 1ml culture media, and disaggregated to single cells as described above. Then 0.5 ml of cells/media suspension was mixed with 2 ml of freezing media and quickly removed to a -80˚C freezer where they were stored overnight before transfer to liquid nitrogen.

3.17.5 Gene Targeting in ES Cells

E14IVtg2a cells were grown on gelatinized plates in culture medium with LIF. $1x10^7$ cells were electroporated (240 V; 500 μF) with a Gene Pulser II (Bio-Rad, Hercules, CA) with 100 μg of targeting vector which had been linearized by *Hind III*. Electroporated cells were grown in culture medium with selection (10 days with G418 (150 μg/ml), after which resistant clones were picked up and expanded. Single clones were picked and grown in 96 well plates. When the cells were confluent, they were split in the following manner: 80 % frozen into liquid nitrogen, 20 % plated into fresh 96 well plates and grown O/N as normal.

3.17.6 Picking Colonies

The plates were washed with PBS. Individual clones were placed in each well of a 96 well "V" bottomed plate (Costar) with 50 μl of trypsin per well and incubated for a few minutes at 37 ˚C. The clones were transferred into a flat bottomed 96 well culture plate (Costar) and grown in culture medium.

3.17.7 Transient Expression of Cre

In order to remove the *loxP* flanked selectable cassette, 10^7 cells of the correctly targeted ES clones were electroporated (settings: 3 μF, 0.8 kV) with 100 μg of *Cre* plasmid. 1000 cells were plated per 10 cm dish. The clones were grown in culture medium without G418.

3.17.8 ES cell transfection

Transfection of ES cell was performed using TransFast Transfection Reagent (Promega, Cat. No. E2341) according to manufacturers instructions. Briefly; LHED targeted cells were plated at densities of $1x10^6$ per 10 cm plate, transiently transfected with 20 μg of the transposase vector pCMV-SB100x (Matessi and Schneider 2009), using transfast and cultured for 48 hours before the addition of

puromycin (2 µg/ml). After 10 days of selection, resistant colonies were picked into 96 well plates, grown and DNA produced for analysis.

3.17.9 Blastocyst Injections

The cell lines to be injected into mice were grown in 24 well plates (Costar). One day before the injection, the cells were washed with PBS and trypsinised into four wells at different plate densities. On the injection day the best well was chosen, depending on the confluency and morphology, and disaggregated to a single cell suspension in culture media. The cells were injected into C57BL/6J blastocysts by Paul S. Devenney (MRC Human Genetics Unit). Briefly, female donor mice (C57BL6/J) were superovulated and set up for matings with C57BL/6J stud males. The following morning they were checked for vaginal plugs, and positive plugs were considered as being 0.5 dpc. Female mice were sacrificed at 2.5 dpc by cervical dislocation and 8-cell embryos and morula were harvested and cultured overnight in M16 medium (Sigma) in a 37 ˚C/5% CO_2 incubator. The following day, blastocysts were selected and injected with 12- 16 ES cells per blastocyst. Injected blastocysts were transferred to 0.5 d or 2.5 d pseudopregnant recipient female mice.

As the ES cells result in chimeric progeny, male mice were set up for test breeding with C57BL/6J female mice to select for germline transmission. Agouti offspring denotes the contribution of the injected ES cells to the germline. Agouti mice were ear clipped and genotyped by PCR.

3.17.10 Tetraploid complementation assay

Cells were prepared as if for blastocyst injection. On the injection day the best well was chosen, depending on the confluency and morphology, and disaggregated to cell suspension including clumps of 4-8 cells in culture media. The cells were injected into C57BL/6J tetraploid blastocysts produced by electrofusion by Paul S. Devenney (MRC Human Genetics Unit). Blastocysts were then transferred to 0.5 d or 2.5 d pseudopregnant recipient female mice.

3.17.11 Retinoic Acid treatment of ES cells

ES cells were seeded at defined densities onto gelatinized plates and grown on normal LIF-containing ES cell media. After 24 hours cells were washed x3 in PBS and media was changed to that lacking LIF and supplemented with 1μM all-trans retinoic acid (Sigma, Cat. No. R2625). At 24 hour intervals media was removed, cells were washed x3 with PBS and fresh media supplemented with retinoic acid was added.

3.17.12 Production of Methaphase nuclei

1 T75 of E14IVtg2a/G4 cells were taken, trypsinised, centrifuged (1000rpm, 5 minutes) and washed in 10 mls ice-cold PBS. Cells were then resuspended in 10 mls hypotonic buffer (KCl, Tri-sodium acetate) and incubated for 10 minutes at room temperature. Again, cells were centrifuged (12000rpm, 4 minutes) and resuspended in 10 mls fixation buffer (3 parts methanol: 1 part acetic acid) and incubated for 10 minutes at room temperature. Cells were washed once more in fixation buffer before being stored at -20 °C in fixation buffer.

3.18 Probe labeling (Biotin)

The following reagents were mixed in order; 2 μl NTS, 2.5 μl of 0.5 mM dATP, 2.5μl of 0.5 mM dCTP, 2.5 μl of 0.5 mM GTP, 2.5 μl bio-16-dNTP, 1 ug template DNA, 1 μl DNAse1 (1:500 dilution), 1 μl DNA polymerase and incubated at 18°C for 90 minutes. After incubation 3 μl 0.5 mM EDTA, 2 μl 20% SDS and 65 μl TE were added to the reaction mix. The reaction mix was then added to a Radiolabeled DNA purification tube (Roche, Cat. No. 11273922001) and centrifuged (2500 rpm, 4 minutes). The eluted DNA was collected in an eppendorf and concentration measured using a Dot blot.

3.19 Fluorescent *in situ* hybridization (FISH)

The following regents were mixed per slide: 80 ng labelled probe, 3 μg Cot-1, 5 μg sonicated salmon sperm plus 2 volumes of ethanol and dried in a spin vac. Pellets were then resuspended in 15 μl hybridization mix (50% deionised formamide, 10% 20x SSC, 19% dH_20, 20% Dextran Sulphate and 1% Tween) and incubated for 1 hour at room temperature.

Two drops of metaphase chromosome solution were added per slide and air-dried for 2 days. Following ageing, slides were washed in 2x SSC, 100 µg RNase at 37°C for 1 hour. Slides were then washed quickly in 2x SSC before being dehydrated through 70%, 90% and 100% EtOH for 2 minutes each and air-dried. Slides were then warmed at 70°C for 5 minutes before being denatured in 70% formamide, 2 x SSC (pH 7.5) at 70°C for 1 minute before dehydration through ice cold 70% EtOH, 90% and 100% EtOH.

Following incubation probes were denatured at 70°C for 5 minutes and pre-annealed for 15 minutes at 37 °C. 10 µl of probe then added to each slide and incubated overnight at 37 °C.

Following incubation slides were washed as follows; 4 x 3 minutes 2 x SSC at 45°C, 4 x 3 minutes 0.1% x SSC at 60°C before transferring to 4 x SSC, 0.1% Tween. Each slide was then incubated with 40 µl blocking buffer (4 x SSC, 5% Marvel) for 5 minutes at room temperature. 40 µl of each antibody then added and incubated at 37°C for 30-60 minutes followed by 3 x 2 minute washes in 4 x SSC, 0.1% Tween at 37°C. Following final wash, slides dried and mounted with 25 µl Vectashield containing DAPI (Vector Labs, Cat. No. H-1200).

3.20 Histology

3.20.1 Whole mount X-gal staining

Whole mount X-gal staining of tetraploids was performed at E11.5. Briefly, E11.5 mouse embryos were dissected into cold PBS, fixed in 4% PFA for 2 hours at 4ºC, these were then washed 3 times in detergent buffer (7 mM Na_2HPO_4, 3 mM NaH_2PO_4, 2 mM $MgCl_2$ ($6H_2O$), 0.1% Na deoxycholate, 0.05% BSA, 0.02% Igepal) and stained overnight at room temperature in staining solution (detergent wash supplemented with 72 mM NaCl, 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$) containing 300 µg/ml X-Gal. After staining embryos were washed in detergent wash, fixed in 4% PFA and stored at 4 ºC.

3.20.2 RNA probe synthesis

The following reagents were mixed in order at room temperature, 1 μl 10x transcription buffer, 1 μl 10x DIG-labelled nucleotide mix, 250 ng template DNA linearised at the 5´ end, 0.25 μl RNAse inhibitor and 1 μl T3 polymerase. The reaction mix was incubated at 37°C for 2 hours; 1 μl of DNAse was added and incubated for a further 15 minutes at 37 °C. 1 μl of EDTA was then added to stop the reaction. RNA purification was performed via the addition of 2.5 μl of lithium chloride and 750 μl anhydrous alcohol and incubation at -20°C for 1 hour. Samples were centrifuged (13,000 rpm; 4 °C; 15 minutes), and the RNA pellet was washed in 500 μl 70% DEPC treated ethanol. A further centrifugation was performed (13,000 rpm; 4°C; 10 minutes). The pellet was air-dried for 5 minutes at room temperature before being resuspended in 50 μl of DEPC treated ddH_2O before being quality and size checked on an agarose gel.

3.20.3 *In situ* hybridisation

CD1 mouse embryos were harvested at embryonic day 11.5 and fixed overnight in 4% paraformaldehyde and stored at -20°C in 100% methanol.

Embryos were rehydrated through a methanol/DepC treated PBT series by washing in 75% MeOH/dPBT, 50% MeOH/dPBT, 25% MeOH/dPBT on ice, each time allowing the embryos to sink. Embryos were then underwent 3x 5 minute washes in dPBT on ice. Embryos were then digested for 20 minutes at room temperature in 10 μg/ml proteinase K in dPBT, before re-fixing in 4% PFA for 45 minutes on ice.

Embryos were pre-hybridised in pre-hyb solution (25 ml Formamide, 12.5 ml 20x dSSC, 1 g blocking powder, 0.5 ml Triton (10%), 2.5 ml CHAPS (10%), 1 ml yeast RNA (50 mg/ml), 0.5 ml EDTA (0.5 M), 250 μl heparin (10 mg/ml) and dH_2O up to 50 ml) by washing twice at room temperature until embryos sank, followed by 1 x 1 hour wash at 65°C and 1x 2-4 hours at 65°C. 5 μl of RNA probe was added to 100 μl pre-hyb solution and denatured at 80°C for 3 minutes, probe was then added to the embryos and incubated overnight at 65°C.

Embryos washed 4x10 minutes at 65°C in 100% post-hyb wash (25 ml formamide, 12.5 ml 20xSSC, 0.5 ml Triton (10%), 2.5 ml CHAPS (10%) and H_2O to 50 ml), 75% Post-hyb/2xSSC, 50% post-hyb/2xSSC and 25% post-hyb/2xSSC, respectively. Embryos were then washed 2x 30 minutes at 65 °C with 2xSSC, 0.1% CHAPS and finally 2x30 minutes at 65 °C with 0.2 x SSC, 0.1% CHAPS.

Embryos were then washed 10 minutes at room temperature in TNT (2.5 ml Tris (1M, pH 7.5), 1.5 ml Nacl (5M), 0.5 ml Triton (10%) and H_2O to 50 ml) before incubation for 4 hours at 4°C in blocking solution (7.5 ml sheep serum (heat inactivated), 1 g BSA, 2.5 ml Tris (1M, pH7.5), 1.5 ml Nacl (5M), 0.5 ml Triton (10%) and H_2O to 50ml). Embryos then incubated overnight at 4 °C in antibody solution (1:2000 α DIG in fresh blocking solution).

Finally, embryos were washed 5x 1 hour at room temperature then incubated overnight at 4 °C in TNT, 0.1% BSA. 2x30 minute washes in NMT, 0-.1% Triton-X-100 (5 ml Tris (1 M, pH9.5), 1 ml NaCl (5 M), 2.5 ml $MgCl_2$ (1 M) and H_2O to 50 ml) followed. Embryos were then washed 3x 10 minutes in NMT before being stained in staining solution (350 μg/ml NBT (Roche Cat. No. 11383213001) and 175 μg/ml BCIP (Cat. No. 11383221001)) in NMT. Finally, embryos were fixed overnight in 4% PFA.

3.20.4 Haematoxylin and Eosin staining

Sections were de-waxed by treating with Xylene (3x5 minutes) before being rehydrated in 3 x 2 minute washes in 100% ethanol followed by 1 x 2 minute wash in each of 90%, 70%, 50% and 30% ethanol before finally being washed in water. Sections were then stained in haematoxylin for 5 minutes and before being differentiated, for a few seconds in 1% HCl in 70% ethanol. A lithium carbonate solution was applied for several seconds to increase the intensity of blue stain before being stained in eosin solution (3 parts 1% aqueous eosin to 1 part 1% ethanol and 0.05% acetic acid) for several minutes. Sections were then rinsed in water and then

100% ethanol for 3 x 2 minutes washes. Finally sections were cleared in 2 x 5 minute wash of xylene and mounted in DPX and cover slip.

3.20.5 Alkaline phosphatase staining

Sections were de-waxed by treating with Xylene (3x10 minutes) before being rehydrated through a series of alcohol washes (3x10 minutes 100%, 1x5 minute 90%, 1x5 minute 70%, 1x 5minute 30%) and rinsed in dH_2O. Following this slides were then washed for 2x5 minutes in NMT (see *in situ* protocol) and stained in staining solution in NMT before being stained in staining solution (3.3 µl/ml NBT and 3.5 µl/ml BCIP) in NMT for 10 minutes. Slides were then fixed for 30 minutes in 4% PFA and mounted using Vectashield (Vector Labs, Cat. No. H-1400).

3.21 Microscopy

For fluorescence imaging, the imaging system comprised a Coolsnap HQ CCD camera (Photometrics Ltd, Tucson, AZ) Zeiss Axioplan II fluorescence microscope with Plan-neofluar objectives, a 100W Hg source (Carl Zeiss, Welwyn Garden City, UK) and Chroma #83000 triple band pass filter set (Chroma Technology Corp., Rockingham, VT) with the excitation filters installed in a motorised filter wheel (Ludl Electronic Products, Hawthorne, NY). For brightfield colour imaging, colour additive filters (Andover Corporation, Salem, NH) installed in a motorised filter wheel (Ludl Electronic Products, Hawthorne, NY) were used sequentially to collect red, green and blue images, that were then superimposed to form a colour image. Image capture and analysis were performed using in-house scripts written for IPLab Spectrum (Scanalytics Corp, Fairfax,VA).

3.21 OPT analysis

Optical projection tomography (OPT) imaging was performed as described by (Sharpe et al. 2002). Briefly, PFA-fixed embryos were embedded in 1% low melting point agarose and then immersed in methanol for 24 hours to remove all water. The embryo was then cleared for 24 hours in BABB (1 part benzyl alcohol/2 parts benzyl benzoate). The sample was then scanned using a Bioptonics 3001 scanner (www.biooptonics.com) with an image taken every 0.9 degrees (of a 360

degree rotation). Upon completion, images are reconstructed using Bioptonics proprietary software with the outputs then being viewed with Data viewer (Bioptonics) and Bioptonics Viewer. Additional 3D outputs were produced using Drishti rendering software (Ajay Limaye-Volume Exploration and Presentation Tool) by Harris Morrison (MRC Human Genetics Unit).

3.22 Mouse lines

The mouse lines used during the course of this study were C57BL/6 (Charles Rivers), CD1 (Charles Rivers), Shh knock-out line described by (Chiang et al. 1996) and a germ line Cre (glcre) produced in house by DJ Kleinjan using a construct previously described by (Araki et al. 1995). A bat *Zrs* 'Knock-in' line created by Laura lettice was also utilised.

All mouse work had ethical approval and was performed under home office legislation using project liscence number (PPL) 60/4424 and personal lisence number (PIL) 60/13367.

Table 3.1: Primer Sequence

Targeting Vector Primers (5'-3')		
Lmbr1	Arm 1(F)	GATCATGCGGCCGC**GGCGCGCC**AAGCAAAGCACCA GCCATT
	(R)	GATCATATCGAT GACCAATGAGCTCCAAGGAT
	Arm 2(F)	GATCATATCGATCCAAGCCTCAGCTGTTCTTC
	(R)	GATCATCTGCAG**TTAATTAA**AGGCAGTGAAGGATTT GGAA
Shh	Arm 1(F)	GATCATGCGGCCGC**TTAATTAA**GACTTCTGGGGAGG GAGAAC
	(R)	GATCATATCGATCAAAATTGGGAATGGTTGG
	Arm 2(F)	GATCATATCGATCCATGGACTTTACTGAGCATTTC
	(R)	GATCATGTCGAC**GGCGCGCC**GCACCTATACTCTGTG TTTTGTGG
Shh2	Arm1(F)	GATCATGCGGCCGC**TTAATTAA**CCACATCCGATAGA

		GGGCTA
	(R)	GATCATATCGATAAGCTTGTTTGCAATCCTACCAGCAA
	Arm2(F)	GATCATATCGATAAGCTTGGGCCTCATCAAGGAATTTT
	(R)	GATCATGTCGAC**GGCGCGCCC**TCATGAGCCCTTCCATACTG
Upstream Lmbr1	Arm1(F)	GATCATGCGGCCGCGCCCATGGAAGAGGTTCCTCAG
	(R)	GATCATATCGATCTCTCAGAAGAATGCTGTTTGG
	Arm2(F)	GATCATATCGATGGATTCAAGTGTGCATGTGG
	(R)	GATCATGTCGACTGGAGTGATAAGGGCAGGAG
Hygro	Arm1(F)	GATCATGCGGCCGCCTGCCGGTAGATTTGAACTCC
	(R)	GATCATATCGATGCTAGCTCTCAGCAGCCTGT
	Arm2(F)	GATCATGTCGACGAGCACAAAAGCCCTGAGAC
	(R)	GATCATCTCGACCTGAAGCCCAATGCTTTCTC
Screening of Colony PCR		
	M13(-20)F	CGCCAGGGTTTTCCCAGTCACGAC
	R	AGCGGATAACAATTTCACACAGGA
	Hygro F	CTTGCTCCTTCGCTTTCTG
	R	TATCCACGCCCTCCTACATC
Screening of ES cell clones for correct targeting		
	M13	TGTAAAACGACGGCCAGTGAGC
	Neo	GCCTTCTATCGCCTTCTTCTTGACG
	SHH F1	GCACACTCATCAGTATCTCACC
	SHH F2 (gap)	GCCATGTTAGTCACATCATGCAC
	SHH R1	GTTTATTGTATGTAGTAAAGTAAG
	SHH R2 (gap)	CATTTACCACATCCCTAGCTGTG
	LMBR1 F1	GACATCTGAACCAGCACCACAG
	LMBR1 F2 (gap)	TCAAGTCTGACTAGGATAACTTAG
	LMBR1 R1	TGCATTATGTGAGTATCAGATCTC
	LMBR1 R2 (gap)	TAGTTTGAGCCTTAGGTGTGCAC
	UpLmbr1 F	GCCCCAAATCAAATCAGAAG
	UpLmbr1 R	TCAGTCCATCATCACCAAGG
Screening of the SB transposon re-insertion		
	Puro: PGK	CGTGGGCTTGTACTCGGTC
		GACAGCACCGCTGAGCAATG
	LacZ(F)	CAACTTAATCGCCTTGCAGCAC
	(R)	CGCTGATTTGTGTAGTCGGTT

Identifying the pLHED re-insertion site		
	SBR1	CTTCTGACCCACTGGGAATGTGATG
	SBR2	GTGGTGATCCTAACTGACCTAAGAC
	SBR3	TCCTAACTGACCTAAGACAGG
	SBL1	CTGGAATTGTGATACAGTGAATTATAAGTG
	SBL2	CTTGTGTCATGCACAAAGTAGATGTCC
	SBL3	AAGTAGATGTCCTAACTGACTTGC
	KmonP-N7-CTCAG	GTACGAGAATCGCTGTCCTNNNNNNNCTCAG
	KmonP-N7-TCCTG	GTACGAGAATCGCTGTCCTNNNNNNNTCCTG
	KmonP	GTACGAGAATCGCTGTCCT
Detecting LHED Deletion		
	Lactopuro	CATAAAATGAATGCAATTGTTG
	ScreenA*	TATTTATAAACCTTATTAAAGCTC
	pLHEDPac*	GAACTGCCAGAACTGTAAGGC
Cloning the *in situ* probes		
	Nom1	GATCATGTCGACGGCTCAGGTTCTGAGACTCG GATCATGCGGCCGCGTCAACACCCTCCGTAGGAA
	Rnf32	GATCATGTCGACTCAGCCATGCCCAATATGTA GATCATGCGGCCGCACAGATGTGCACAGGACAGG,
	Rbm33	GATCATGTCGACGGGCAACACTTGAGACCATT GATCATGCGGCCGCCTGGACATCAGTGGTGGATG,
	Lmbr1	GATCATGTCGACCTGTGATGTCCAGAACACTG GATCATGCGGCCGCGATTCTGTAAGATGAATCAG.
	Mnx1	GATCATGTCGACTGGACACCCGGTCTACAGTT GATCATGCGGCCGCCCATTGCTGTACGGGAAGTT
RT-PCR analysis		
	Lmbr1	GCTGGTTGATGAGACTGCAA GTGCTTTCTGATGCCCATTT
	Rbm33	GGGCAACACTTGAGACCATT CTGGACATCAGTGGTGGATG
	Rnf32	TCAGCCATGCCCAATATGTA ACAGATGTGCACAGGACAGG
	Nom1	GGCTCAGGTTCTGAGACTCG GTCAACACCCTCCGTAGGAA
	Shh	GCCTACAAGCAGTTTATTCCCAAC CAGTGGATGTGAGCTTTGGATTC
	Mnx1	TGGACACCCGGTCTACAGTT CCATTGCTGTACGGGAAGTT
	Oct4	GGCGTTCTCTTTGGAAAGGTG CTCGAACCACATCCTTCTCT
	B-actin	GGCCCAGAGCAAGAGAGGTATCC ACGCACGATTTCCCTCTCAGC
Genotyping		
	LacZ	CAACTTAATCGCCTTGCAGCAC CGCTGATTTGTGTAGTCGGTT
	Hygro	GATGTAGGAGGGCGTGGATA GATGTTGGCGACCTCGTATT

	Neo	GATATTCGGCAACCACCCATC TGTTCCGGCTGTCAGCGCAC
	Cre	AGCTGGCCCAAATGTTGCTG CAATTTACTGACCGTACAC
	LHED deletion 5’	CATAAAATGAATGCAATTGTTG TATTTATAAACCTTATTAAAGCTC
	LHED deletion 3’	CATAAAATGAATGCAATTGTTG GAACTGCCAGAACTGTAAGGC
	LHED homozygote	CCACCTGCTTCTCTGAGGAA TATCTCACAGTTGGCCCCAG
	Rnf32	GGGATGCCAAATTAAGGAGA GGGACCAGAAAGTGTATGGC

Bold/underlined sequence represents restriction digest sites; NotI SalI, ClaI PacI and Asc1. *Primers ScreenA or pLHEDPac were used alongside Lactopuro depending on whether deletion was upstream or downstream of the LHED insertion site**.**

4. Results

4.1 Introduction

Until recently tools that allow the survey of large genomic regions from a few kb to a Mb have been limited. However, recently two studies have exploited the Sleeping Beauty transposon for this purpose (Kokubu et al. 2009; Ruf et al. 2011). Sleeping Beauty is a binary system consisting of transposon DNA and a transposase enzyme, which were genetically reconstructed from inactive elements found within the salmon genome. The transposon is flanked by a binding site termed an inverted repeat/direct repeat (IR/DR) whereby the transposase binds and can be used for mobilisation. The SB transposon also has a predisposition for 'local hopping' meaning it provides a means of saturating a genomic region for study. One such system which encompasses Sleeping Beauty for genomic engineering is the LHED system which has previously been discussed (Section 2.16) (Kokubu et al 2009). The system allows the generation of nested deletions at defined loci using *cre* technology as well as a reporter vector which can be used to monitor enhancer activity within a region. The system is performed within ES cells which can then be used to generate mouse lines or using tetraploid complementation embryos can be generated wholly from ES cells. By combining these two techniques a map of enhancer activity can be generated and loss of function phenotypes examined (Figure 4.1).

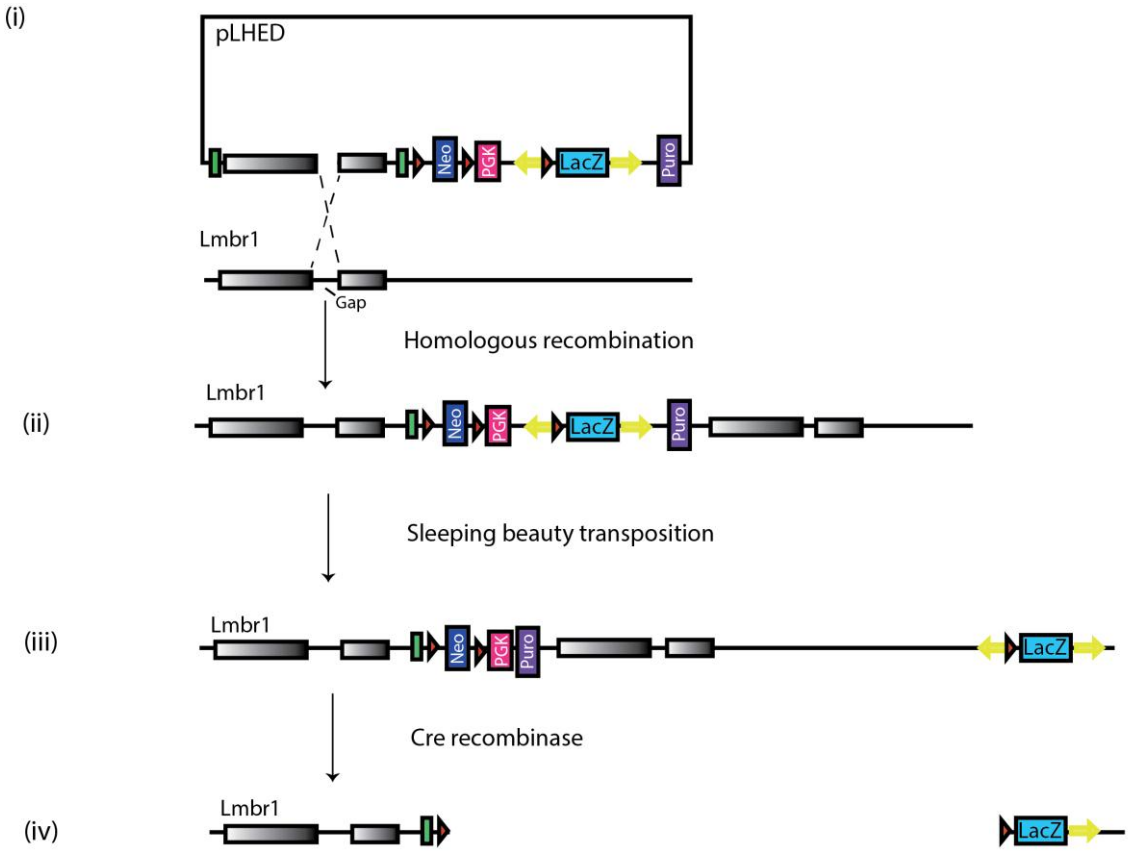


Figure 4.1: The LHED transposon system

(i) The LHED system comprises the pLHED vector containing a *LacZ* reporter gene (light blue rectangle), IR/DR repeats (Yellow arrows), and G418 selection cassette with promoter (dark blue rectangle), puromycin selection cassette (pink rectangle), PGK promoter (purple rectangle), *LoxP* sites (red triangle) and FRT sites (green rectangles). Firstly a homology arm for region of interest complete with pre-positioned gap was cloned into the vector. The vector was then linearised within the gap and targeted into the genomic locus via homologous recombination which can be selected for using neomycin (ii). (iii) Once inserted within the genome the transposon elements consisting of the *LacZ* gene flanked by the SB IR/DR can be mobilised by transient expression of a transposase. Following transposition the PGK promoter comes in contact with the puromycin resistance gene meaning puromycin can be used to select for transposition. The transposon element will then reinsert within the genome within approximately 50% of cases and the *LacZ* can then be mobilised to detecting enhancer function. (iv) *Cre recombinase* can also be utilised to generate deletions between the *LoxP* sites of the original insertion site and the transposed site which can then be utilised to detect loss of function phenotypes. The *LacZ* gene also remains present so alterations to enhancer activity can also be analysed.

Genome regulatory organization mapping with integrated transposons or GROMIT provides a similar technique. It again relies on the mobilisation of the Sleeping Beauty transposon; however, it does not require the sophisticated

manipulations of embryonic stem cells and is instead performed entirely within mice (Ruf et al. 2011). Both systems are similar in that insertions can be remobilised however it has been reported that the GROMIT system provides a higher frequency of remobilisation of one or two orders of magnitude higher than LHED (Ruf et al. 2011). There are several other proposed differences between the systems. Firstly; GROMIT can be utilised to generate chromosomal rearrangements by recombining between chromosomes, not just deletions or inversions like the LHED system (Ruf et al. 2011). Secondly, GROMIT is performed entirely in mice, which is proposed to be more cost and time efficient. However, the generation, maintenance of mouse lines is still very time costly, whereas the use of ES cells means that only clones with insertions of interest need be generated, which makes the system less labour-intensive. Furthermore it is well known that ES cells are capable of supporting the development of the entire embryo and the adult mouse when the trophoblast and primitive endoderm lineages are provided by tetraploid carrier embryos (tetraploid embryo complementation assay) (George et al. 2007). This means that by using ES cells to generate tetraploid embryos, a quick method of looking at enhancer activity can be provided.

The LHED system was therefore chosen as a method of manipulating the *Shh* genomic locus with an aim of gaining a further understanding with regards to enhancer activity and also to generate deletions within the region to see how this affects phenotype.

4.2 Designing the LHED strategy

Firstly the LHED vector was designed to allow targeting into the region surrounding *Shh* and the ZRS.

4.2.1 Designing homology arms for exploration of the Shh genomic region

The LHED vector was designed and constructed to contain no *BamHI* or *HindIII* restriction enzyme sites, thus allowing the use of any genomic region encompassing a *BamHI* or *HindIII* site to be chosen as a homology arm containing a unique restriction site (Kokubu et al. 2009).

With regards homology arm design it was proposed preferable to choose a region around 5-7 kb containing a large region of fairly unique DNA in order to increase the chance of homologous recombination. As the gene desert is comprised of large amounts of repetitive DNA it was deemed best to try and design vectors within the *Lmbr1* gene. It was decided that the best locations for the initial insertion vectors were 3´ of the *Zrs* so that resulting remobilisations would allow deletions between that and *Shh* and also 5´ of the *Zrs* so that subsequent remobilisations would allow deletions upstream of the enhancer.

Two pLHED vectors were designed and constructed to insert either side of the *Zrs*; the first comprising a 4.8 kb genomic region 774 kb from *Shh* designated the Shh_LHED vector, and the second a 7.6 kb region located approximately 855 kb from *Shh*, designated the Lmbr1_LHED vector. Both homology arms included a prepositioned gap of 634 bp and 600 bp respectively since this is considered to improve recombination efficiency.

As the Shh_LHED vector later proved to be problematic for carrying out targeting, a further pLHED vector was designed to allow exploration of the *Shh* genomic region, comprising a 7.5 kb region located adjacent to the *Zrs* (781 kb from *Shh*), this vector was designed as a replacement for the Shh_LHED vector and thus was designated Shh2_LHED. Again a prepositioned gap was engineered comprising 666 bp (Figure 4.2.).

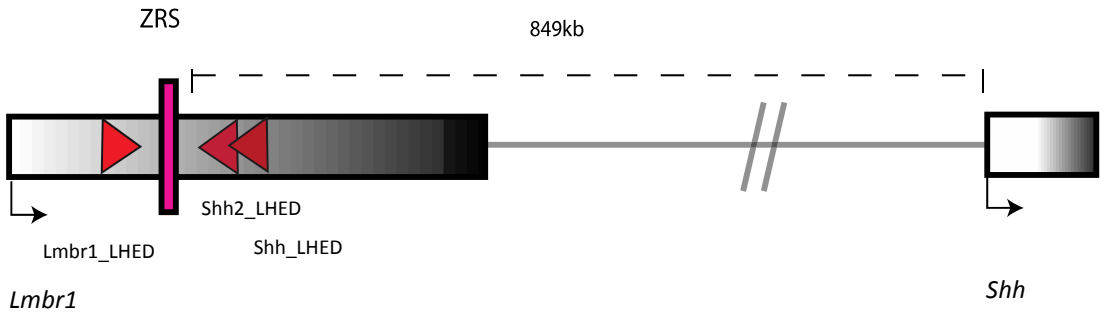


Figure 4.2: LHED targeting into the *Shh* region

Three LHED vectors (red arrows) were designed to insert either side of the Zrs (pink rectangle) within the *Lmbr1* gene (grey rectangle) these were at a distance of ~800Kb from the *Shh* promoter. The orientation of the vectors was of importance for utilisation of the *loxP* sites.

4.2.2 PCR amplification and pBluescript cloning

Often cloning large vectors is time-consuming and limited by the need for appropriate restriction enzyme sites. A simpler approach for generating vectors for ES cell targeting is homologous recombination via a process known as gap repair, which provides a convenient method for subcloning DNA from BACS into pBluescipt (Liu et al. 2003).

The first stage of the cloning process was to design PCR primers which were then used to amplify 200-500 bp regions of the BAC (PAC 542-N10), which ultimately represent the ends of the fragment to be sub cloned by gap repair. Although smaller fragments can be used it has been shown that larger homology arms significantly increase the frequency of sub cloning by gap repair (Liu, Jenkins, and Copeland 2003).

Primers were designed to amplify 300 bp regions from the 5´ and 3´ ends of each of the homology arms (Lmbr1, Shh and Shh2) from BAC DNA encompassing the *Lmbr1* genomic region (See table 3.1 Targeting vector primers for further details). Primers were also designed to include restriction enzyme sites firstly for cloning into pBluescript and then additionally for cloning and positioning the homology arm in the correct orientation in the pLHED vector. Correct-sized bands were visualized on a gel before being digested using appropriate restriction enzymes

(Figure 4.3 A). Ligations were then performed to clone each end separately into the vector pBluescript which had also been digested with the appropriate enzymes. Following transformation, small-scale plasmid purification was performed on a select number of colonies and the resulting DNA was sequenced in order to confirm that no mutations had been induced during the PCR for any of the clones (Using M13 primers see Table 3.1 for further details). The vectors containing the 3´ end of the homology arm was then digested with restriction enzymes (*NotI/ClaI*) and a ligation set up to ligate the 5´ end of the homology arm into the same vector. Following transformation a colony PCR was used to identify vectors containing both ends of the homology arm. These positive colonies were then purified using small scale plasmid purification and a diagnostic digest was used to confirm that an insert representing both ends of the homology arm were present (Figure 4.3 B).

4.2.3 Bacterial recombineering

A gap repair method was utilised in order to sub clone the entire homology arm fragment from BAC DNA. This was achieved by generating double strand break within the pBluescript vector containing the homology arm ends. The DNA was subsequently electroporated into electro-competent EL350 gap repair cells carrying the relevant BAC, which had been induced for the recombination genes *exo*, *bet* and *gam* expression by prior growth at 42 °C. Resulting colonies were purified by small scale plasmid purification and a diagnostic digest was used to confirm gap repair had occurred and a complete homology arm was present (Figure 4.3C).

The homology arm fragments were then sub-cloned into the pLHED vector in several clones. This was confirmed via restriction digest (Figure 4.3 D).

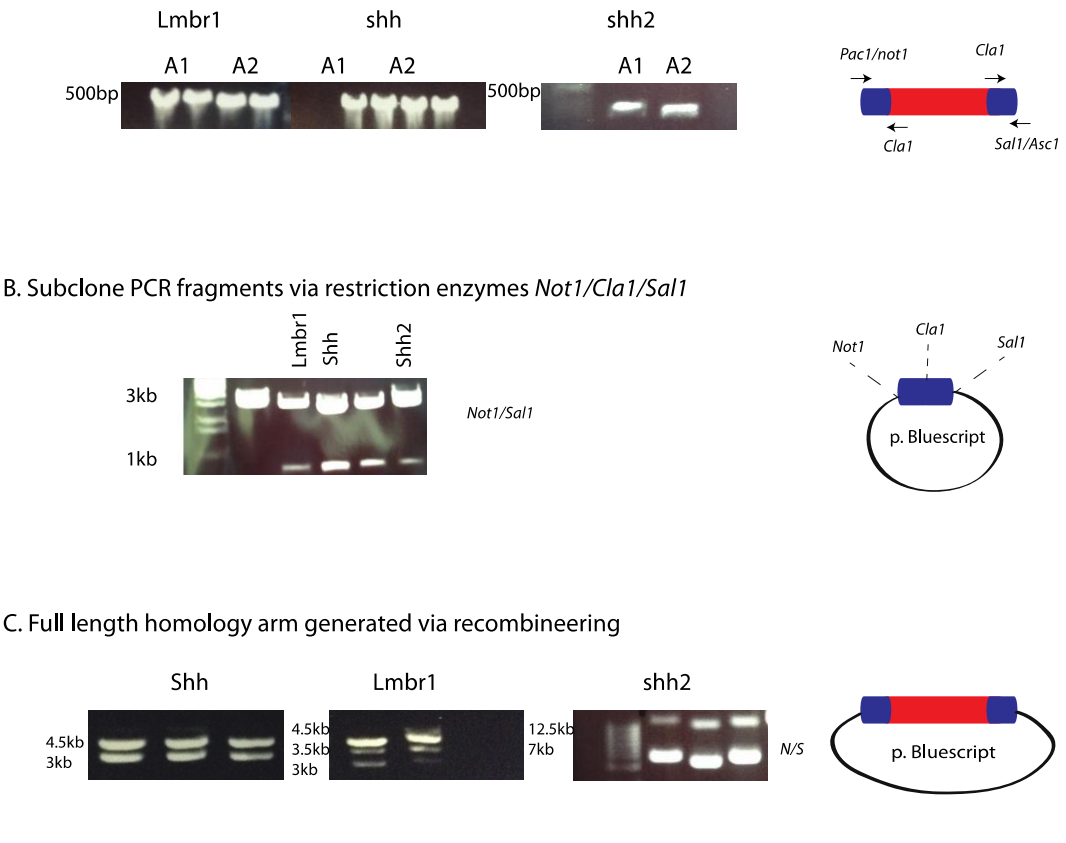


Figure 4.3: Homology Arm Cloning

A. 300 bp fragments from either side of the three homology arms; Lmbr1_LHED, Shh_LHED and Shh2_LHED were generated by PCR. Primers were designed to include restriction sites to enable cloning into vectors.

B. The PCR fragments were then cloned together into the pBluescript vector by utilising the *Not/Sal* restriction sites and a restriction digest was used to show the 600 bp fragment had dropped out from the 3 kb vector.

C. The p.Bluescript vector was linearised via digest using *Cla1* and recombineering was used to generate the full homology arm of each of the vector. This was then confirmed by restriction digest using *Not1* and *Sal1*.

D. The full length homology arm was then sub cloned into the 12.5 kb pLHED vector which was confirmed by restriction digest. Multiple bands were present for each of the Shh_LHED vectors as a result of additional *Pac1* sites.

4.3 G4 ES cell targeting

Both Lmbr1_LHED and Shh_LHED vectors were linearised via a restriction enzyme site within the gap and electroporated into G4 ES cells. The G4 ES cell line is an F1 hybrid ES cell line from a 129 and C57BL/6 cross (George et al. 2007). The use of hybrid ES cell lines has previously been shown to have increased frequency for producing completely ES cell derived embryos with the G4 line in particular having been shown to support tetraploid embryos to term at frequency of 70% (George et al. 2007) The transformants were then selected using G418 (the pLHED vector carries neomycin resistance). Following 10 days selection approximately 100 colonies were isolated for each electroporation, these clonal cells were then expanded, lysed and DNA was extracted.

4.3.1 PCR analysis of targeted clones

The next step within the targeting process was to determine which clones had been correctly targeted with the pLHED vector into the correct genomic loci. The pLHED vector is targeted into a genomic region using insertion type homologous recombination. During this process the pre-positioned gap has been shown to be efficiently repaired using the chromosomal target sequence (Valancius and Smithies 1991) and the homology arms are duplicated (Figure 4.4A). The gap therefore presents as a unique region of DNA which should only be present alongside the vector if targeting has occurred within the correct genomic region.

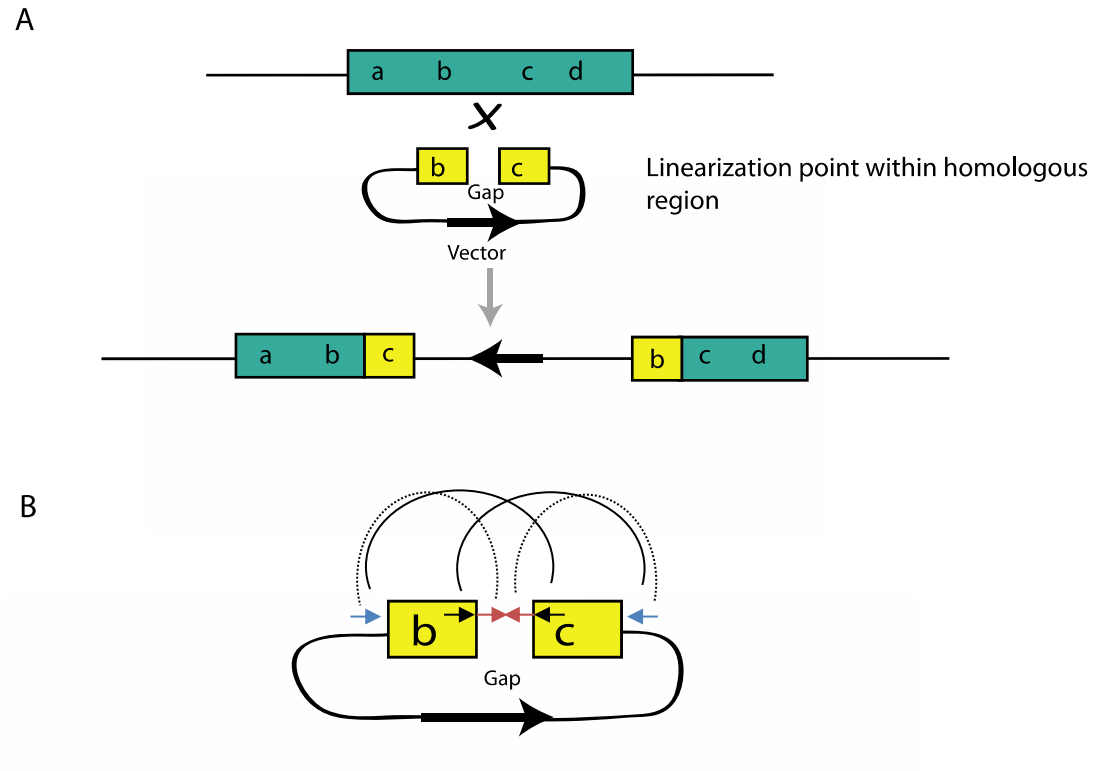


Figure 4.4: Homologous recombination and PCR design

A. The insertion vector method. Vectors are linearised within a pre-positioned gap within the homologous region. Sequences within the locus are duplicated, and marker is inserted at the homologous site.
B. PCR primers were designed against vector sequence (shown with blue arrows) and also within the homology arm (black arrows) which would amplification over the gap region. An additional set of primers were also designed within the gap region (red arrow) which represents a unique region of DNA which should only repaired if the vector has inserted within the correct location. Primer details are given within Table 3.1 'Screening of ES cell clones for correct targeting'; M13/Neo representing vector specific primers, SHH/LMBR1 F/R representing those which amplified over the gap and SHH/LMBR1 F2/R2 representing those present within the gap.

PCR primers were designed corresponding to sequence within the gap region of the homology arm and each end of the vector sequence (Table 3.1 'Screening of ES cell clones for correct targeting'), thus allowing identification of clones which had experienced homology arm duplication, gap repair and contained vector backbone (Figure 4.4B). PCR analysis was performed using these primers for both targeted cell lines however it was not possible to identify any clones positive for this PCR. A second PCR was therefore devised in which primers flanking the gap rather than within the gap were used (Figure 4.4B) and identified four clones which were

positive for insertion of the Lmbr1_LHED vector and homology arm duplication (Figure 4.5A). It was not possible to identify any Shh_LHED positive clones

The resulting Lmbr1_LHED positive clones were retrieved from long term storage and grown for several days before being used to generate metaphase chromosomes in order to examine the karyotype. Within the cell lines in which PCR suggested correct targeting more than 70% of cells were found to have undergone massive chromosomal duplications and rearrangements. This was a phenomenon found only to be associated with the G4 ES cell line suggesting that these cells are more susceptible to chromosomal rearrangements. This number of abnormal cells was deemed too high to be used to generate mouse lines therefore it was decided that these cells could be used no further

4.4 E14 ES cell targeting

It was decided that the rearrangements found to have occurred within the targeted G4 cells were an artefact of the specific strain of cells rather than the targeting therefore a new series of targeting was undertaken using E14TG2a ES cells (E14), again using both the Lmbr1_LHED and Shh_LHED targeting vectors. E14 ES cells are not a hybrid line but instead derived entirely from 129 animals. They therefore do not have the same efficiency for supporting tetraploid embryos to term but they can support them until at least E11.5 and they can also be used for blastocyst injection to generate chimeric animals for mouse lines (Yang et al. 2009). The vectors were linearised and electroporated into E14 ES cells before being selected using G418, following 10 days selection, approximately 100 colonies were isolated and DNA extraction was performed.

4.4.1 PCR analysis of E14 targeted ES cells

PCR analysis was performed on both the Lmbr1_pLHED and Shh_pLHED targeted E14 ES cell clones to determine which clones had been correctly targeted. Long-range PCR analysis was undertaken using two sets of primers, one

corresponding to vector backbone and gap DNA sequence and a second again corresponding to vector sequence but instead of using primers against the gap sequence, primers flanking the gap were utilised (Figure 4.4B/Table 3.1). Approximately 10% (10/98) of clones transfected with the Lmbr1_LHED vector were found to be positive for correct targeting as deemed by the primers flanking the gap (Figure 4.5B), however only 8% (7/98) were identified by primers within the gap (Figure 4.5D). With regards to the Shh_LHED vector, again approximately 10% (7/90) of clones were believed to be positive for targeting using primers flanking the gap (Figure 4.5C) however when primers within the gap were utilised it was not possible to identify any Shh_LHED clones in which correct targeting had occurred.

As a result of the failed attempts to target the Shh_LHED vector it was decided that a third vector, the Shh2_LHED vector would be used to target E14 ES cells. The second Shh vector (Shh2_LHED) was homologous to a similar region as the first Shh vector (Shh_LHED), however, had longer homology arms and contained a slightly higher percentage of repetitive non-unique DNA. Approximately 100 clones were screened by long-range PCR using primers homologous to LHED vector sequence and the gap region of the homology arms were then used to screen ES clones. Two positively targeted clones were identified (2%) by PCR (Figure 4.5D). The fact that it was possible to target the second vector and not the first therefore suggests that a minimum homology arm length is critical for the ability to target cells.

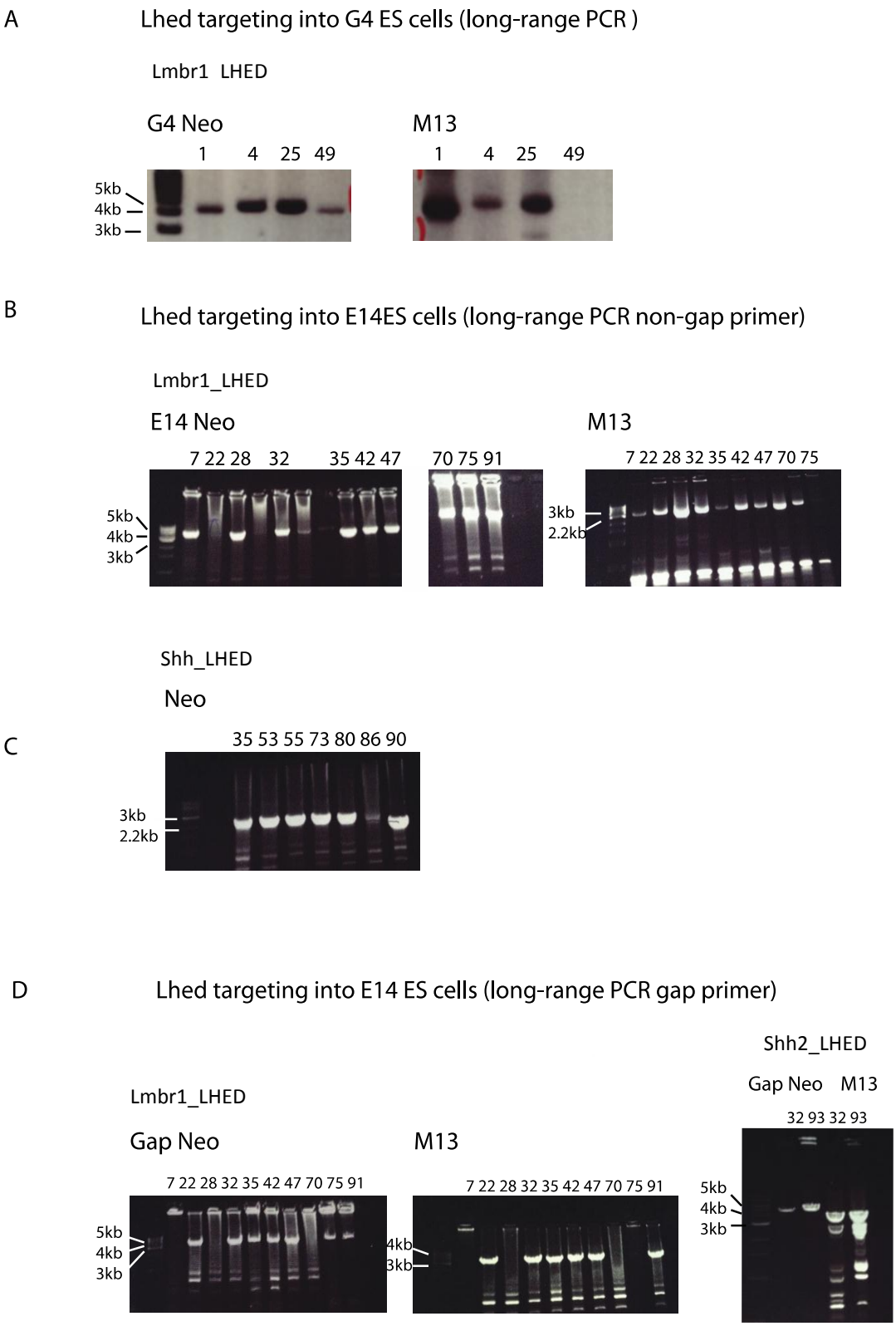


Figure 4.5: ES cell targeting

A. PCR analysis of Lmbr1_LHED targeting into G4 ES cells. Long-range PCR analysis undertaken from both end of the vector (Neo versus M13) and primers flanking the gap. Positive clones were indicated by a positive PCR band at 4 kb and 3.5 kb respectively.
B. PCR analysis of Lmbr1_LHED targeting into E14 ES cells. Long-range PCR analysis undertaken from both end of the vector (Neo versus M13) and primers flanking the gap. Positive clones were indicated by a positive PCR band at 4 kb and 3.5 kb respectively.
C. PCR analysis of Shh_LHED targeting into E14 ES cells. Long-range PCR analysis undertaken from one end of the vector (Neo) and primers flanking the gap. Positive clones were indicated by a positive PCR band at 2.5kb.
D. PCR analysis of Lmbr1_LHED targeting into E14 ES cells. Long-range PCR analysis undertaken from both end of the vector (Neo versus M13) and primers within the gap. Positive clones were indicated by a positive PCR band at 4kb and 3.5kb respectively. PCR analysis of Shh2_LHED targeting into E14 ES cells. Long-range PCR analysis undertaken from both end of the vector (Neo versus M13) and primers within the gap. Positive clones were indicated by a positive PCR band at 5 kb and 4 kb respectively.

4.5 FISH analysis of correctly targeted ES cells

In order to confirm that the LHED vectors had been correctly targeted into the genomic region surrounding the *Zrs*, fluorescent *in situ* hybridisation (FISH) was employed. Probes were generated for both the LHED vector backbone and also the *Rnf32* gene (located adjacent to *Lmbr1* on chromosome 5) thus co-localisation of the probes would indicate that the vector had targeted into the correct genomic region. Positive clones were retrieved from liquid nitrogen storage and grown before being used to generate metaphase chromosomes for FISH. Probes were indeed found to co-localise for clones in which gap primers suggested correct targeting had occurred (Figure 4.6 B) (N=3), however random integration was found for those in which only flanking primers had yielded a positive PCR (Figure 4.6A) (N=3).

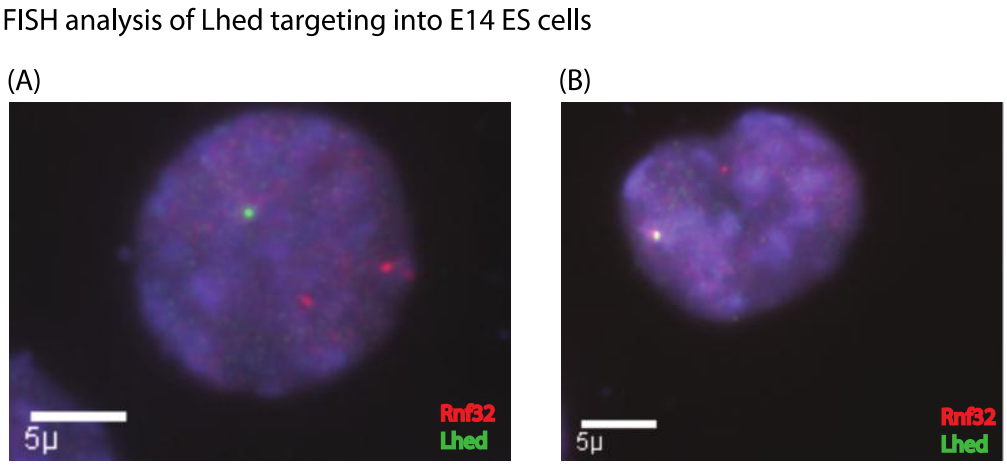


Figure 4.6: FISH analysis of correctly targeted ES cell lines

FISH analysis of ES cells targeted with LHED_Lmbr1 vector. Green probe represents pLHED vector and the red probe the *Rnf32* gene.

(A) Probes not co-localised suggesting random integration.

(B) Probes co-localised suggesting correct targeting on chromosome 5.

4.6 Inducing transposition in targeted ES cells

Following targeting of the LHED vectors Lmbr1_LHED and Shh2_LHED into the *Shh* locus the next step was to mobilize the SB transposon cassette using a transient transfection of transposase. The Sleeping Beauty transposase binds to the terminal inverted repeats/direct repeats present within the transposon causing it as well as the *LacZ* reporter gene and LoxP site contained within them to 'jump'. This jumping brings together a puromycin selection cassette with a PGK promoter meaning that puromycin resistance can be used to determine in which clones hopping has occurred (Figure 4.1 [iii]).

Targeted cells were electroporated with the CMV-SBII transposase plasmid in order to mobilize the transposable element. Initially, puromycin selection at a concentration of 1 µg/ml was employed after 24 hours to identify clones in which the vector had mobilized. Unfortunately all cells were found to die following selection; this was hypothesised to be as a result of the time required following electroporation for the transposase to be expressed, the transposon to jump and the puromycin to be expressed. The electroporation was therefore repeated with recovery times of 48 and

72 hours before selection was administered. However, in these experiments a further problem was identified in that a large number of background cells were still present after selection, suggesting some background puromycin resistance. Puromycin kill curves were performed on both targeted cell lines as well as wild-type E14 ES cells. 100% of wild-type cells were found to die at puromycin concentrations of 1µg/ml after 48 hours selection; however under the same conditions approximately 5-10% of targeted cells were still alive. Whereas, puromycin concentrations of 2 µg/ml were found sufficient to cause 100% cell death within both targeted cell lines (Table 4.1). The puromycin resistance found within the targeted E14 ES cells was likely as a result of the vector being 'leaky'. This phenomenon did not occur in the original study however a different cell line was used (G4 rather than E14) which may account for this. Electroporation was therefore repeated and selected using the higher concentration of Puromycin after 48 hours; unfortunately no clones were produced suggesting there was a fundamental problem with the LHED system.

Table 4.1: Inducing hopping with LHED targeted ES cells

Cell death analysed following puromycin treatment for 48 hours on Wild-type E14 ES cells, Cell lines L35 and L47 positively transfected with Lmbr1_LHED vector and cell line S32 transfected with Shh2_LHED vector at concentrations of 1µg/ml and 2µg/ml, a minimum N=3 applied for each condition.

Cell Line	Average % cells alive following 48 hour treatment with 1µg/ml Puromycin	% cells alive following 48 hour treatment with 2µg/ml Puromycin
Wild type E14	0	0
L35	5	0
L47	5	0
S32	5	0

In order to determine what was preventing the transposition event occurring two further aspects of the experimental procedure were tested. Firstly, the transfection method was examined, electroporation had been used to transfect the cells previously; however, it was hypothesised that a reagent based transfection method might be more efficient. The second aspect of the experiment that was tested was the transposase, the transposase initially used was the SBII transposase as discussed in the original LHED paper (Kokubu et al. 2009); however, a newer version of the Sleeping Beauty transposase, SB100x has been produced. SB100x has been hyper-mutated by including six combinational units from TC1 Mariner to increase its efficiency 100 fold over the original transposase (Matessi and Schneider 2009). Transfection with the SB100x transposase (and a chemical transfection reagent) was therefore tested and was found to be capable of producing clones (Summary Table 4.2).

Table 4.2: Summary of transfections used to optimise transposase activation within LHED target cell lines

	Cell line	**Number of plates and cell density per plate**	**Transposase used and concentration of transposase per plate**	**Transfection method**	**Selection time point**	**Concentration of Puromycin**	**Number of Clones**	**Further information**
1	L35	10 x plates, 1x 10^6 cells per plate	SB11 10 µg/µl	Electroporation	24 hours	1 µg/ml	0	
2	L35	12 x plates, 0.8 x 10^6 cells per plate	SB11 20 µg/µl	Electroporation	48 hours (4 plates), 72 hours (4 plates), 96 hours (4 plates)	1 µg/ml	8 clones present across the 12 plates. 6 of which on 72 hour plates	Further PCR an analysis suggested all to be lacZ positive
3	L35	10 plates, 1x 10^6 cells per plate	SB11 20 µg/µl	Electroporation	72 hours	1 µg/ml	0	Large number of background cells (Puro resistant) prevented analysis of clone number
4	L35	30 plates, 1x 10^6 cells per plate	SB11 20 µg/µl	Electroporation	72 hours	1 µg/ml	0	Large number of background cells as above
5	L35	10 plates, 1x 10^6 cells per plate	SB11 20 µg/µl	Electroporation	48 hours	2 µg/ml	0	
6	L47	10 plates, 1x 10^6 cells per plate	SB11 10 µg/µl	Electroporation	48 hours	2 µg/ml	0	

7	S32	10 plates, 1x 10^6 cells per plate	SB11 10 µg/µl	Electroporation	48 hours	2 µg/ml	0	
8	S32	7 plates, 1x 10^6 cells per plate	SB11 10 µg/µl	Transfection reagent	48 hours	2 µg/ml	0	
9	S32	7 plates, 1x 10^6 cells per plate	SB100x 10 µg/µl	Transfection reagent	48 hours	2 µg/ml	250	200 clones analysed for reinsertion site
10	S32	8 plates, 1x 10^6 cells per plate	SB100x 10 µg/µl	Transfection reagent	48 hours	2 µg/ml	250	202 clones analysed for reinsertion site
11	L35	8 plates, 1x 10^6 cells per plate	SB100x 10 µg/µl	Transfection reagent	48 hours	2 µg/ml	160	139 clones analysed for reinsertion site

4.7 Determining 'Hopping' and Reinsertion within LHED_SHH2 Targeted cell Lines

Two transfections were performed using the Shh2_LHED targeted cell line (designated S32) and SB100X, each producing approximately 250 clones. DNA was extracted from these for analysis and a series of PCRs employed. Firstly, in order to detect that clones had definitely 'hopped' a PCR was performed using primers for the PGK promoter and Puromycin gene (See Table 3.1 'Screening of SB transposon site'), these two elements come together after transposition, therefore, a positive PCR indicated that the transposition had occurred. All clones tested were found to be positive for transposition. Following 'hopping' the transposon, was reported in the literature to reinsert in approximately 50% of the clones (Kokubu et al. 2009). In order to determine which clones had reinsertion of the LHED transposon PCR was performed using primers against the *LacZ* reporter gene (See Table 3.1 'Screening of SB transposon site'). This gene is found within the mobilised section of the LHED vector (Figure 4.1 [iii]) therefore a positive PCR suggests that the LHED vector had reinserted within the genome. Transposon reinsertion was detected in approximately 50% of clones. Finally, in order to detect the location of reinsertion a nested asymmetric PCR strategy was employed, which had previously shown to be capable of mapping transposon insertions generated by the GROMIT system (Ruf et al.

2011). Briefly, the strategy involved an initial PCR using primers facing out either side of the transposon sequence and a random primer which was designed to amplify approximately every 1000 bp, therefore meaning that it should be possible to generate a PCR product for a high percentage of insertions regardless of where the transposon had inserted. A second PCR was then performed again with transposon-specific primers and also with primers against an adaptor located within the random primers (For primer details see Table 3.1 'Identifying the pLHED reinsertion site'). This second PCR was then sequenced to determine where the transposon had inserted (See appendix table 8.1). Of the 200 clones positive for reinsertion, it was possible to generate long enough sequence to unequivocally map 86%. Of these 86%, 65% were found to map within the same chromosome (chromosome 5). With 41% of these found to map to within 1Mb of the initial insertion site (Table 4.3/Figure 4.8).

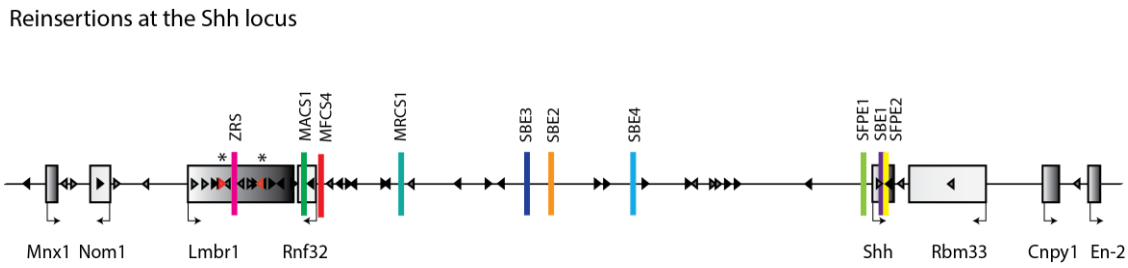


Figure 4.7: Mapping reinsertions at the Shh locus in LHED-targeted ES cells

The *Shh* regulatory landscape containing the genes *Mnx1*, *Nom1*, *Lmbr1*, *Rnf32*, *Shh*, *Rbm33*, *Cnpy1*, *En2* (grey rectangles-Dark 5'-3' expressing, Light 3'-5') and various labelled enhancers (coloured rectangles).
Reinsertions of the LHED targeting vector from both original insertion sites; S32 and L35 (red arrowheads demarked with *) were mapped via a nested high throughput PCR strategy, and their locations and orientation then plotted throughout the region (black arrow heads reflecting reinsertions in S32 line and grey triangles reflecting reinsertion within L35 line with the direction of the arrowhead reflecting orientation of *LacZ* transcription). Large amount of coverage across the region found.

4.8 Determining 'Hopping' and Reinsertion within lmbr1_LHED Targeted cell Lines

One transfection using the SB100x transposase was performed for Lmbr1_LHED targeted cell line (designated L35) and was found to produce 165 clones. Again these clones were analysed using the PCR strategy discussed

previously. Approximately 60% of clones were found positive for reinsertion, with the reinsertion site of these unequivocally mapped for 93% (See appendix table 8.1). 58% of reinsertions were found to occur within chromosome 5, with 46% of these within 1 Mb of the targeted insertion site (Table 4.3/Figure 4.8).

The original study (Kokubu et al., 2009) suggested that approximately 78% of cells have a reinsertion within the same chromosome which is slightly higher than found within these experiments. However 42% of insertions were found within a 1 Mb region which is similar to that found within these experiments. Again in agreement with the original study, an approximate 50:50 ratio was observed with respect to which orientation the vector was found to reinsert in for both vectors. Reinsertions were also found to cluster around the original targeting site (Table 4.3) an event which is likely as a result of the local hopping nature of the sleeping beauty transposon.

Table 4.3: Distribution of LHED insertions across the mouse genome and within chromosome 5

Majority of insertions shown to be on chromosome 5; with the biggest number of these between 20 and 30Mbp* which is also the region of the original insertion and thus demonstrates the propensity of the SB transposon for local hopping, the number of insertions appears to tail of either side.

Chromosome	No. of insertions		Location on chromosome 5 (Mbp)	No. of insertions
1	5		0-10	4
2	4		10-20	14
3	5		20-30*	81
4	7		30-40	14
5	158		40-50	12
6	3		50-60	12
7	8		60-70	7
8	3		70-80	5
9	4		80-90	0

10	6		90-100	5
11	10		100-110	0
12	4		110-120	1
13	5		120-130	2
14	6		130-140	1
15	4		140-150	0
16	1			
17	3			
18	3			
19	3			
x	2			

4.9 Conclusions

Two LHED vectors Shh2_LHED and Lmbr1_LHED, were successfully cloned and used for ES cell targeting. Originally it was planned that ES cell targeting would be performed using G4 ES cells which have previously been shown to be superior for tetraploid production (George et al. 2007). Unfortunately successful targeting of these cells was not found to be possible. Following PCR analysis, karyotype analysis was undertaken on those clones which were expected to be correctly targeted. However a number of chromosome rearrangements and duplications were found to have occurred. Other members of the lab using these cells had experienced similar problems, even with cells that had not been through a round of targeting. This potentially suggests a fundamental issue with the original stock of cells or alternatively as a result of trypsin sensitivity. Further work was performed in order to examine this. A round of targeting was employed on G4 ES cells cultured without trypsin, using instead TryplE express (Invitrogen) and appeared to have a normal karyotype, however, as a great deal of work had already been performed using the E14 ES cells it was decided that it was more beneficial to carry on using these instead.

Developing a PCR screen capable of detecting correct insertions was required. The original PCR screen in which primers were designed corresponding to gap and vector sequence was not found to work within the first round of targeting into G4 cells. This was hypothesized to be as a result of incomplete gap repair, therefore a new PCR screen that involved amplifying across the gap was developed. It was believed that this PCR screen would still be able to identify correctly targeted clones as it detects the presence of homology arm duplication which occurs during homologous recombination into the region. The same PCR screens were used to analyse clones produced by targeting into E14 ES cells, both screens were found to detect positive clones, however, only those clones in which PCR suggested gap repair had occurred were found by FISH to have inserted into the correct genomic region. This not only confirms that gap repair does occur during targeting reactions but also that it is an essential indicator of targeting into the correct genomic region. The clones that had not undergone gap repair were found to have inserted at random into the genome. The band sizes generated by PCR analysis on clones which had inserted at random suggested that the homology arm had been duplicated, however as these vectors were not found within the desired locus it appears that a semi illegitimate recombination reaction has occurred, potentially in which concatemers of vector had inserted. Often events such as these occur if the targeting vector has not been fully linearised (Valancius and Smithies 1991). PCR analysis of the Shh_LHED vector using primers across the gap suggested that approximately 10% of clones had been correctly targeted. However FISH analysis suggested that these clones had undergone random integration. It is therefore suggested that it was not possible to target this vector correctly. The most obvious reason for why this has happened is that the Shh_LHED vector has a slightly smaller region of homology, which has previously been shown to affect targeting frequency (Valancius and Smithies 1991).

Experimental optimisation with regards transfection method, puromycin selection concentration and transposase had to be employed in order to induce transposition within the LHED cell lines. Electroporation was originally employed because it had previously been demonstrated within the lab to have a reasonable efficiency for transfecting ES cells. However it was suggested that the transposase

step works more efficiently if a reagent-based transfection method was used (personal correspondence Chikara Kokubu). There was also an issue with regards to underlying puromycin resistance within non-transposed cells. All cells carrying the LHED vector also carry a puromycin resistance gene however without its promoter. It is therefore likely that the cause of this increased puromycin resistance is a result of low level puromycin resistance transcript being produced. This problem was shown to be remedied by increasing the concentration of the selection reagent without any detrimental effect on the cells. Finally a different transposase was tested; SB100X which is believed to have a 100x greater efficiency than SB11 which was employed within the original study. It was indeed found possible to induce LHED transposon mobilisation using SB100x rather than SB11. The fact that no clones were produced using the SB11 transposase vector suggests that there could be a fundamental problem with the vector being utilised.

5. Mapping the *Shh* regulatory region

5.1 Introduction

As previously described, the *Shh* gene and its well-established regulators provides an example of a genomic landscape in which enhancers reside in a large desert extending into neighbouring genes to control the spatiotemporal pattern of expression. This composition coordinates a complicated series of control mechanisms posing a number of questions about the capacity of the regulatory components within these large chromatin domains. Various enhancers have been identified within the region which directs expression within the CNS, gut epithelium as well as the limb bud (See Figure 2.3). A library of LHED insertions carrying the *LacZ* reporter gene were dispersed throughout the *Shh* region via targeting of the two LHED targeting vectors and subsequent mobilisation of the transposable element (Figure 4.8). These insertions thus provided a means to map enhancer activity throughout the *Shh* regulatory domain to determine a potential relationship between enhancer activity and its position within the locus and elucidate on a chromosomal scale the regulatory events displayed over a long range.

The advent of powerful genomic technologies has allowed for the unprecedented genome-wide prediction of enhancer elements (Visel, Rubin, et al. 2009; Rada-Iglesias and Wysocka 2011). However, to rapidly validate and characterize the vast number of genomic regions with regulatory potential remains a considerable challenge, particularly in mammals. In a typical enhancer assay, a putative enhancer element is placed upstream of a minimal promoter driving a reporter gene the expression of which is analysed in an appropriate experimental system (Loots 2008). There are significant drawbacks associated with many of the existing methodologies. For example, transient transfections of reporter constructs are easy to perform but assess the elements in multiple copies and in a non-chromosomal context. At the other end of the spectrum, mouse transgenesis offers a more natural *in vivo* context, however it is expensive, time-consuming, and laborious as well being a skilled technique not always widely available.

ES cell tetraploid complementation embryos provide a method of generating mice entirely derived from ES cells. This is achieved by removing two-cell stage embryos from the oviducts of superovulated females and treating them with a short electric pulse in order to fuse two blastomeres. The tetraploid embryos then have their *zonae pellucidae* removed and are aggregated with clumps of 8-15 ES cells before being transferred into the uterus of pseudopregant female mice (Nagy et al. 1993). This produces embryos derived entirely from ES cells. Embryos can then be harvested at an appropriate stage and examined.

5.2 Generation of tetraploid embryos

In order to analyse *cis*-regulatory reporter function, it is sufficient to generate heterozygous embryos carrying an LHED insertion. A number of ES cell clones carrying SBLac inserts (comprising a *LacZ* reporter gene) were selected which were dispersed throughout the *Shh* genomic locus, as well as both upstream and downstream. These were then injected into tetraploid blastocysts which were subsequently used to generate embryos derived wholly from ES cells. At E11.5 whole mounts of the resulting embryos were analysed for enhancer activity by examining LacZ expression from the transposed transgene.

5.3 Reporter activity across the gene desert

The first region of the *Shh* regulatory landscape to be examined was the gene desert in which a number of the *Shh* enhancers reside. ES clones were chosen to examine the activity of reporter genes between *Shh* and the *Rnf32* gene. The gene desert covers nearly 800 kb of DNA containing a number of well-characterised *Shh* enhancers (Previously described in section 2.8). Four ES cell clones were selected in order to examine enhancer activity across the gene desert starting within just outside the *Rnf32* gene (SBLac 2-120) and ending 100 kb from the *Shh* gene (SBLac 120) (Figure 5.1).

All four reporter insertions showed similar expression patterns of *LacZ* reporter gene activity (Figure 5.1). Expression was found to occur within the brain, notochord and floor plate, throughout the lungs and gut and also within the posterior of the limb bud. This reflected the known *Shh* expression pattern as shown by *in situ* hybridisation (Figure 5.7). Enhancer activity was found to be similar regardless of where the insertion was placed within the region thus suggesting that the whole desert is open for enhancer activity (Figure 5.1).

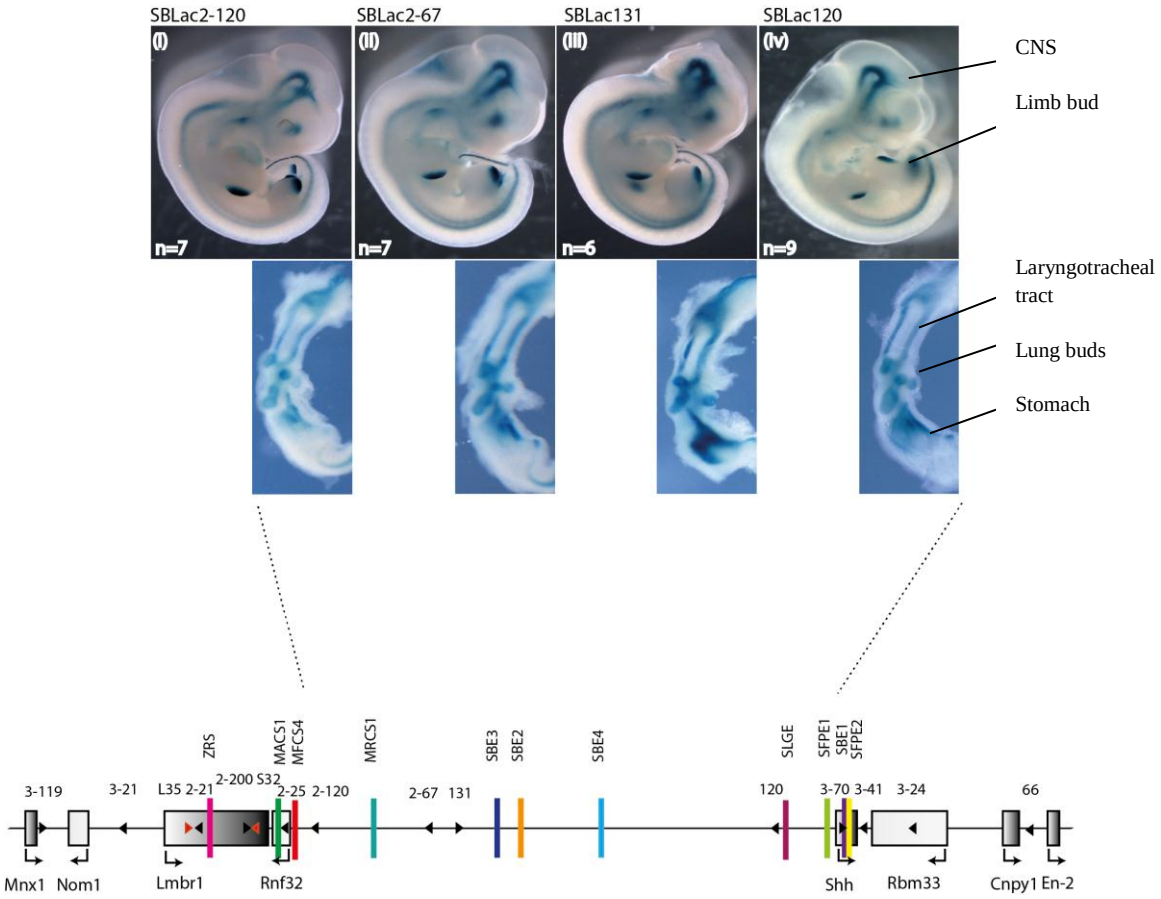


Figure 5.1: Enhancer activity within the gene desert

Expression of *LacZ* at different locations across the gene desert. LacZ is detected within the CNS, the limb bud and the throughout the gut including the laryngotracheal tract. Whole embryo and gut expression from four insertions within the gene desert (i) 2-120 (ii) 2-67 (iii) 131 (iv) 120. Insertions indicated by arrow heads on the gene desert map and enhancers shown as coloured rectangles. n numbers reflecting number of embryos examined also shown.

Although tissue expression was found to be similar across the region, some variations were found between insertion sites with regards to levels of expression.

The expression levels found from the SBLac2-120 insertion, especially in the gut epithelial lining appeared to be lower than those seen within the other SBLac inserts (Figure 5.1 [i]). Reporter expression levels within the limb and brain remained similar suggesting this was a phenomenon associated with the gut enhancers and the region the reporter gene had inserted. Enhancers known to direct *Shh* expression within the gut, pharynx and oral epithelium at this stage have been identified and are found approximately 600-900 kb from *Shh* (Sagai et al., 2009). Additionally the *SLGE* enhancer lies close by the SBLac120 insertion (Tsukiji, Amano, and Shiroishi 2013). This phenomenon therefore could result from the compounded effect of both the *MACS1* enhancer and *SLGE* driving expression most strongly towards the *Shh* gene. Expression within the laryngotracheal tract has been suggested to be driven entirely by *MACS1*, whereas both *MACS1* and *SLGE* drive expression throughout the rest of the gut (Tsukiji, Amano, and Shiroishi 2013). Expression within the laryngotracheal tract of the 2-120 tetraploid (Figure 5.1 [i]) was found to mimic levels found within the rest of the gut suggesting that it is the activity of *MACS1* that the reporter gene is detecting. This therefore could suggest that *MACS1* while capable of working at long distances is less effective as an enhancer at short range, further quantification however would have to be performed.

5.4 Do Insertions vary in their intensity of staining?

Following staining overnight it was deemed that the majority of tetraploids had either reached saturation or reached the end of the reaction, therefore by examining embryos only in this manner it was not possible to detect all changes in the intensity of staining or with regards temporal staining patterns i.e. did certain insertion stain faster than others. A time-course experiment was therefore performed whereby several insertions were used to generate tetraploid complementation embryos or if mouse lines had been generated for the same insertion (see chapter 6), F1 embryos. These were harvested at 11.5 days and were then stained for LacZ expression, with the staining process being stopped at 1 hour, 2 hours, 4 hours and then lastly overnight. Four insertions were examined: SBLac120, SBLac2-67, SBLac2-120 and SBLac 2-200 (Figure 5.2). Overall the staining pattern was found to

be linear; after 1 hour staining faint LacZ expression was found within the limb buds and also the midbrain of all the embryos. After 2 hours' staining, the rest of the brain tissues were found to start expressing LacZ and there was also faint expression within the notochord and floor plate. Following 4 hours' staining, LacZ expression appeared to be present within all the known *Shh* expressing tissues of the limb, CNS and gut suggesting that it was possible to detect a complete pattern after this time period. Following staining overnight, the pattern appeared to be saturated. Each of the insertions resulted in some variation in staining which concurred with that identified using the tetraploid complementation assay, for example SBLac 2-120 was slightly lighter than SBLac120 and SBLac 2-67 after the overnight time point, and SBLac2-200 had higher limb bud staining and was missing forebrain expression. Although SBLac2-120 and SBLac2-67 were found to have lower limb and mid-brain expression at the early time points, staining was present suggesting that there were no significant changes in temporal expression.

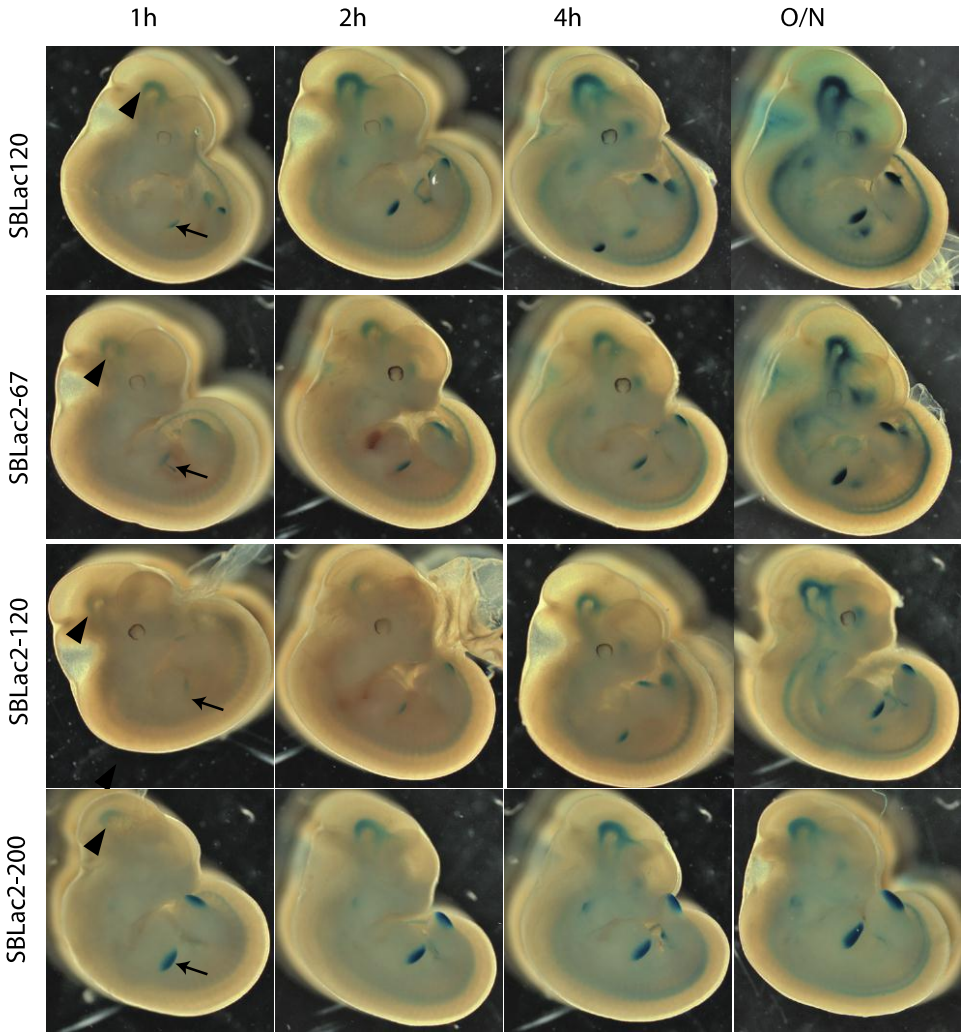


Figure 5.2: Enhancer activity time course

Time course of *LacZ* expression at four time points 1 hr, 2 hr, 4 hr and overnight, for four insertions, 120, 2-67, 2-120 and 2-200. Black arrows mark limb expression at 1 hour time point while black arrow heads show midbrain expression. Expression is shown to begin to be detected within limbs and mid-brain at 1 hour, with the entire expression pattern recapitulated by 4 hours. Enhancer activity lost within some tissues depending on insertion sites suggesting variation in robustness.

5.5 Reporter Gene Activity inside Genes

As well as containing a number of enhancer elements the *Shh* genomic region also contains both the *Rnf32* and *Lmbr1* genes. Both *Lmbr1* and *Rnf32* are believed to be expressed ubiquitously throughout the embryo however no developmental role has been defined for either. The *Rnf32* promoter has been shown to be surrounded by enhancers related to *Shh* expression (See section 2.8) however its ubiquitous

expression suggests it is not influenced by these. To determine if there is an intrinsic property of the *Rnf32* gene that contributes to the lack of responsiveness to the regulatory landscape, an insertion, SBLac2-25, which is located with an intron of the gene, was selected to generate tetraploid embryos for expression analysis. SBLac2-25 was found to be expressed throughout the CNS, brain, gut and limb in a typical *Shh* spatial pattern (Figure 5.3B [V]). The reporter expression levels, however, in these embryos were consistently lower than the reporters in the gene desert, except perhaps for the limb buds. This suggests that the *Rnf32* which resides inside the regulatory landscape and a promoter which is active is resistant to outside enhancer influence; however, the body of the gene is accessible to the surrounding expression activity suggesting that the transcriptional apparatus for *Rnf32* expression is refractory to outside influence.

The *Lmbr1* gene shows a similar resistance to the *Shh* regulatory landscape; however as its promoter is outside the regulatory region this could explain why it is not influenced by enhancers within the region. The two initial targeting sites of the SBLac vector; SL35 and S32 both found within *Lmbr1* (either side of the *Zrs*) were originally chosen to generate tetraploid embryos to examine reporter activity within this region (Figure 5.3). The overall expression pattern was found to be low for both insertions; however, in the SBLacS32 the expression pattern appeared to be complete except for lack of expression within the ventral forebrain and gut (Figure 5.3A [ii]). A similar exclusion of the ventral forebrain was found in the SBLacL35 insertion; in addition, expression in the neural tube and notochord were very low and expression within the gut was missing (Figure 5.3A[i]).

Prior to transposition (Figure 4.1), the *LacZ* reporter gene remains with the rest of the targeting vector which contains a number of other promoters for both the neomycin and puromycin resistance genes, the presence of these in such close proximity to the *lacZ* reporter gene could dilute down the amount of enhancer activity focused on its own promoter. As there were concerns about comparing original insertion sites versus those which had 'jumped', two further insertions within *Lmbr1* were examined for expression, one either side of the *Zrs*; SBLac2-21 and SBlac2-200 which were comparable to the original two in terms of location but in the opposite orientation and had also undergone transposition. The expression

pattern within the SBLac2-200 insertion was again found to be more or less complete, however lower levels of expression were found within the ventral forebrain, notochord and gut (Figure 5.3B [iv]). In SBLac2-21 ventral forebrain expression was lost as was the expression within the gut, expression in the neural tube and notochord was also found to be quite low (Figure 5.3B [iii]). Both tetraploid embryos generated within this region were found to have very high limb expression which could be as a result of the proximity of the reporter to the *Zrs*. Overall the expression pattern found within tetraploid embryos generated from the original insertion sites versus those generated from transposed insertions within the *Lmbr1* gene were found to be comparable. The tetraploids generated post-excision were found to have on the whole a somewhat higher level of LacZ expression, potentially as a result of not being in close proximity to the remaining vector sequence. There did not appear to be any effect of orientation of the insertion on the expression pattern.

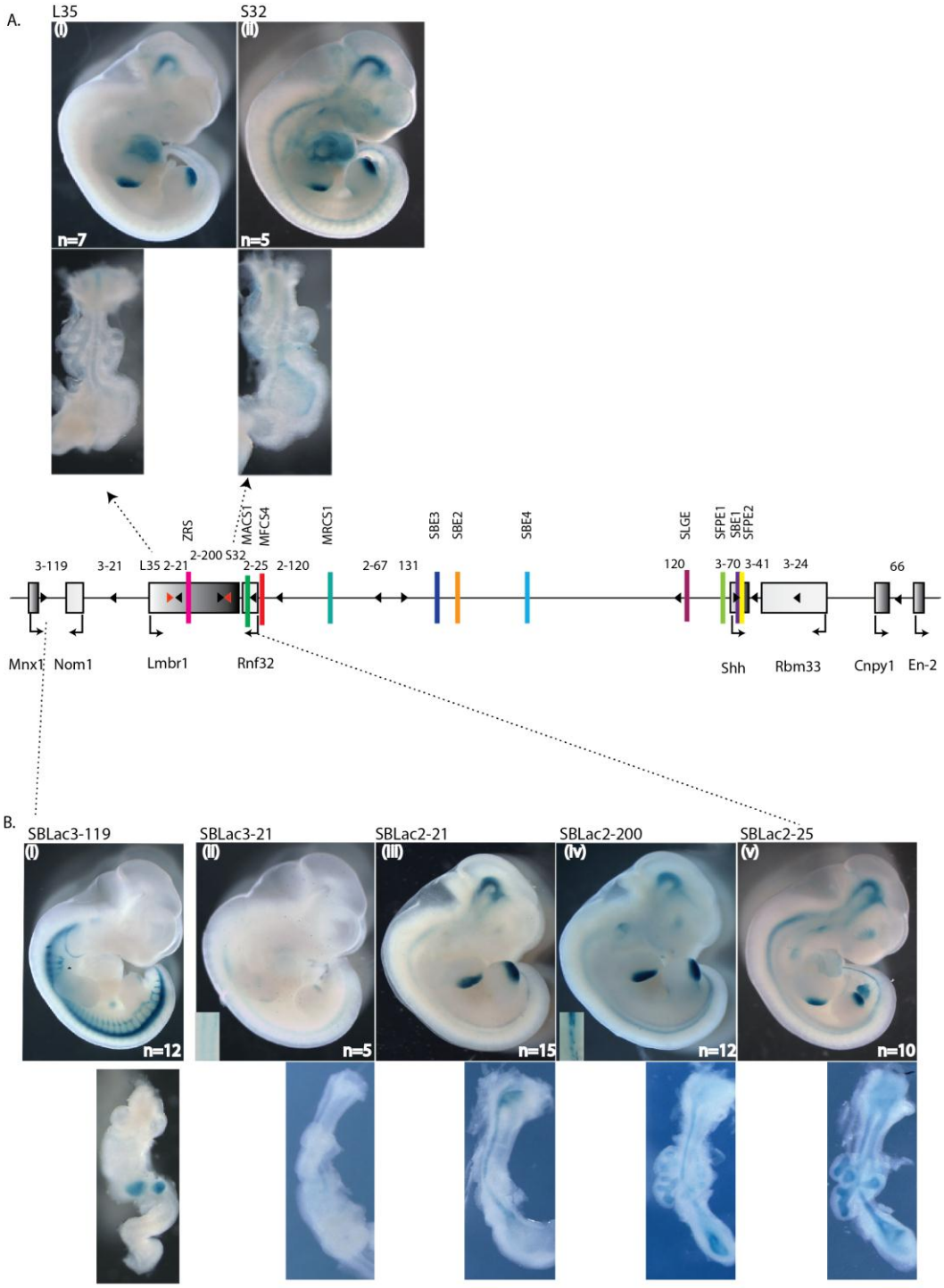


Figure 5.3: Enhancer activity at the 5' end of the *Shh* regulatory landscape

A. Expression of *lacZ* reporter gene with two original insertions (i) L35 and (ii) S32 indicated on the gene map by red arrow heads. Expression is missing from the forebrains of both embryos (black arrow), additionally only low level gut expression is found.

B. Expression variation at the 5' end of the *Shh* regulatory region. Whole embryo and gut expression from five insertions upstream of *Lmbr1*, within the *Lmbr1* gene and within *Rnf32* (i) 3-119 (ii) 3-21 (iii) 2-21 (iv) 2-200 (v) 2-25. Neural tube dissection shown at bottom left hand corner for SBLac3-21 and SBLac2-200. Insertions indicated by arrow heads on the gene desert

map and enhancers shown as coloured rectangles. N numbers shown for number of tetraploid embryos generated for each insertion.

Another insertion within a gene, SBLac70, located at the other end of the region within *Shh* was used to examine reporter gene expression pattern. Tetraploid complementation embryos generated for this insertion were found to have very high expression levels within tissues, especially the mid-brain and also the gut (Figure 5.4 [i]). The expression pattern seen within embryos generated from the SBLac3-70 insertion could suggest a focusing of enhancers within the region of *Shh*. Alternatively the known mid-brain enhancer SBE1 resides within the *Shh* gene, which could suggest proximity to the enhancer is the reason for the high degree of expression seen within these tissues. However as the gut is also shown to have a very high expression pattern, transcriptional read-through could be occurring due to the orientation of the insertion within the gene in the same direction as the gene itself, resulting in very high levels of expression.

5.6 Defining the boundaries of *Shh* enhancer activity

A series of tetraploid complements were generated to examine the boundaries of the *Shh* region. Upstream of *Lmbr1*, two tetraploid complements were generated, SBLac3-21 and SBLac3-119. In the former line, the reporter locates approximately 10 kb from the 5´ end of *Lmbr1* and in the latter, 100 kb from *Lmbr1* past the next gene, *Nom1*. Both insertions were found to show some LacZ expression however in a pattern dissimilar from the known *Shh* expression. Within SBLac3-21 expression was exhibited both in the neural tube and also in the limb (Figure 5.4 B [ii]). SBLac 3-119 was also shown to express LacZ within the motor neurons and faintly within the hind limb (Figure 5.4 B [i]). No *Shh*-related expression was found within the guts of embryos generated from these insertions. These embryos were dissected and sectioned in order to detect if the neural tube expression seen was related to *Shh*. The neural tube of the SBLac3-21 embryos was found to have an expression pattern unlikely to be *Shh* related, with expression found in two stripes within the dorsal root

ganglia region rather than within the floor plate and notochord as seen with Shh expression. Limb expression from these two insertions was also found to be slightly altered as opposed to the *Shh* related tetraploids with expression only present within the posterior hind limbs.

Three insertions were also examined downstream of *Shh*, firstly SBLac3-41 located within the intergenic region between *Shh* and the nearest neighbouring gene *Rbm33* approximately 5 kb from *Shh*. Secondly, an insertion within an intronic region of the *Rbm33* gene SBLac24, 100 kb from *Shh* and lastly an insertion between *Cnpy1* and *En-2*, SBLac1-66. The expression pattern of SBLac3-41 was shown to reflect that of the *Shh* pattern within the CNS, gut and limbs. However expression levels were much lower than those seen within *Shh* and also the other side of *Shh* (Figure 5.4 [ii]). The following insertion SBlac24 located within the *Rbm33* gene showed no *Shh*-related expression, instead expression was found within the hind brain-mid brain junction (Figure 5.4[iii]). A similar expression pattern was found for the final insert SBLac1-66, thus suggesting that these insertions lay within another regulatory domain separate from *Shh* (Figure 5.4 (iv)).

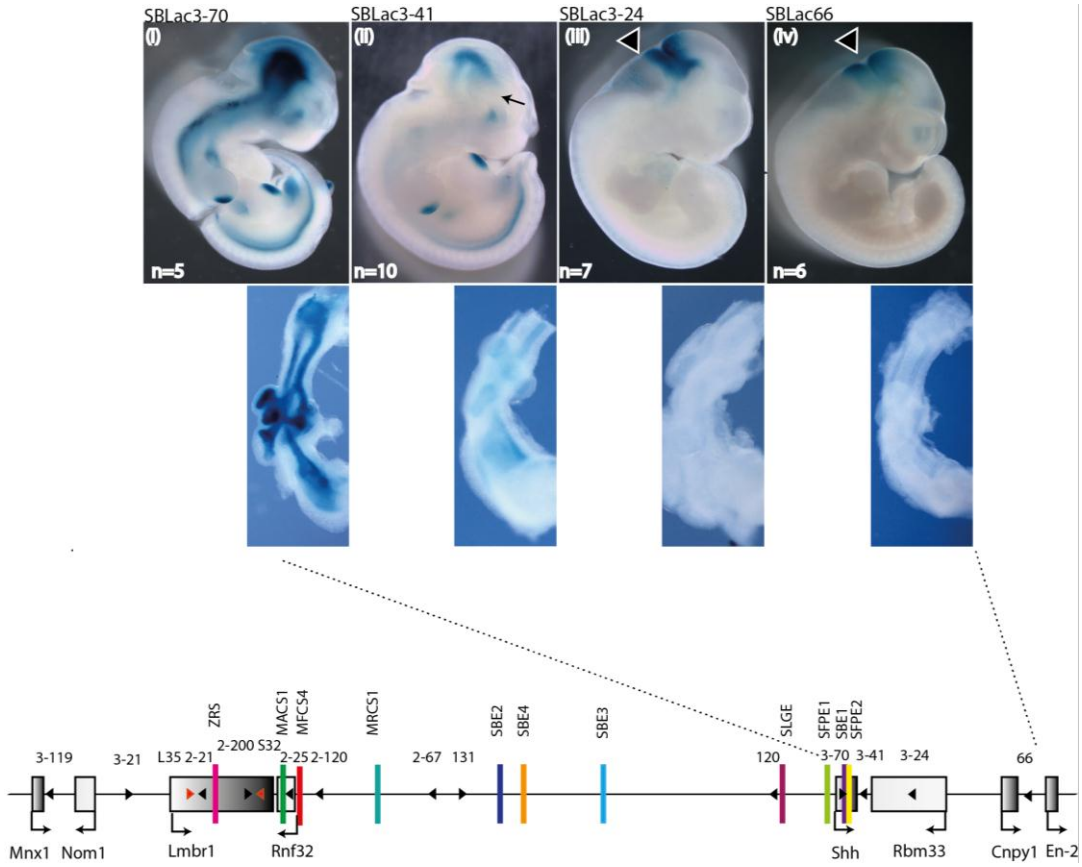


Figure 5.4: Enhancer activity at the 3' end of the *Shh* regulatory landscape

Expression variation at the 3´ end of the *Shh* regulatory region. Whole embryo and gut expression from three insertions downstream of *Shh* and within the *Shh* gene and (i) 3-70 (ii) 3-41 (iii) 3-24 (iv) 1-66. Black arrow used to show loss of forebrain expression within SBLac 3-41 whereas black arrow heads highlight expression within mid-brain hindbrain junction within SBlac3-24 and SBLac 1-66. Insertions indicated by arrow heads on the gene desert map and enhancers shown as coloured rectangles

5.7 How robust are enhancers within the *Shh* regulatory region?

Overall the enhancers found within the *Shh* genomic region appeared to have fairly consistent expression patterns and activity across the region. However some variations were seen with regards how robust single enhancers were.

Optical Projection tomography (OPT) analysis was used in order to the compare whole embryo *LacZ* expression patterns for several different insertions. OPT allows visualisation of tissues throughout the embryo that cannot necessarily be

visualised by microscopy alone. Four embryos were selected for analysis, SBLac120 located 100 kb from *Shh*, SBLac2-67 within the middle of the region, SBLac3-41 located 10 kb downstream of *Shh* and SBLac2-21 located at the 5' end of *Lmbr1*. Embryos were found to show similar expression patterns, however, some differences were found. Both SBLac120 and SBLac67 were found to show the full pattern of expression within the CNS, gut and limb (Figure 5.5 [ii] [iii]). However the SBLac120 insertion was found to be expressed more highly in most embryonic sites (Figure 5.5 [ii] [iii]). SBLac3-41 was found to show expression within the mid-brain, some parts of the notochord and also within some of the gut and the posterior of the limb bud, however at lower levels (Figure 5.5 [iv]). Brain, limb and a little notochord expression was exhibited within SBLac2-21, with gut expression and fore-brain expression found to be missing (Figure 5.5 [i]). This data goes to show us that although the whole region of the desert is open for reporter activation there are some variation in how enhancers work. The three enhancers *SBE2*, *SBE3* and *SBE4* located within the middle of the region which control expression within the forebrain and telencephalon are found to be the most sensitive to changes in distance. Whereas, *SBE1* which is found within intron 2 of the *Shh* gene and responsible for driving expression within the mid-brain from E10.0 onwards is found capable of acting at a very long distance over 700 kb away. This again demonstrated the variation in enhancer activities but also goes to reveal that enhancer proximity to the promoter is not necessarily critical for function.

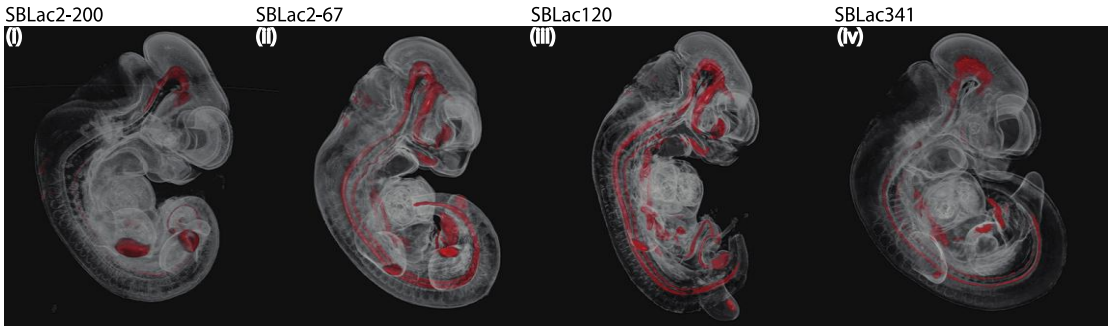


Figure 5.5: OPT analysis of insertions across the *Shh* region

OPT images showing sagittal view of four separate insertions (i) 2-200 (ii) 2-67 (iii) 120 and (iv) 3-41. LacZ expression pattern is represented in red with the structures of the embryo shown in grey. SBLac120 is shown to recapitulate the entire *Shh* expression pattern, with the other insertions showing loss of expression within some structures suggesting variation in enhancer robustness.

5.8 Does the transposon vector carry any genetic information when it jumps?

A concern with using a transposon based method of chromosomal engineering was the possibility that its mobilisation would result in indels within the region. In order to evaluate whether this had occurred sequence from 65 insertion sites was examined, of these it was possible to deduce that within at least 80% of insertions no changes had occurred to sequence after insertion, with the number potentially being higher if sequence quality had been higher. When the transposon mobilised only the sequence located between the inverted/direct repeats is supposed to 'hop' (Figure 4.1). However as tetraploid embryos across the region were shown to have very similar expression pattern it was decided that two insertions out with the *Shh* regulatory region should be examined in order to ensure that no genetic information from the original insertion locus was carried with the transposon when it jumped. Two separate inserts; SBLac 189 and SBLac 2-89 which are both located on chromosome 5; however at distances of 12 and 23 Mb from *Shh*, respectively, were used to generate tetraploid complementation embryos. SBLac189 was found to yield no staining whatsoever, suggesting it had landed within a region of the genome with no enhancers in range or it had landed within a region of closed chromatin (Figure 5.6 [i]). SBLac 2-89 was found to have some staining at the top of the mid-brain and also within some regions of the lungs and gut (Figure 5.6 [ii]). However this pattern was very different from the one demonstrating *Shh* type expression. Both these insertions go to show that no information from the original insertion site was carried with the transposon when it jumped.

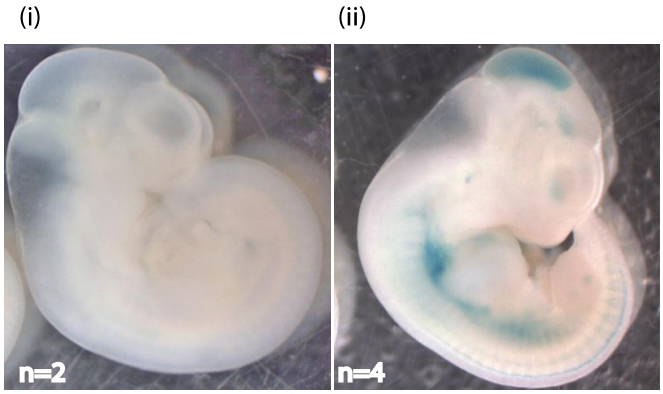


Figure 5.6: LHED transposition does not result in insertions within the genomic region

Tetraploids generated from two insertions located elsewhere on chromosome 5 (i) 189 and (ii) 2-88. SBLac189 shows no expression pattern suggesting no enhancer is functioning within range of the insertion. SBLac2-89 shows expression within structures of the gut and the top of the mid-brain. A different expression pattern to that of *Shh*.

5.9 What are the expression patterns of neighbouring genes within the *Shh* region?

A number of genes are located within, and in close proximity to the *Shh* regulatory locus. It was of interest whether any of these genes are influenced by any of the enhancers located within the region.

5.9.1 *In situ* hybridisation probe cloning

In order to evaluate how the genes both within and in close proximity to the *Shh* region are expressed an *In situ* hybridisation screen was undertaken. Primers were designed to amplify approximately 500 bp regions spanning at least one exon from whole embryo cDNA (Figure 5.7A[i]) (Table 3.1 'Cloning *in situ* probes'). Sequencing and restriction digest was then used to ensure the correct sequences had been cloned (Figure 5.7 A[ii]). *In situ* hybridisation was performed on E11.5 embryos in order to determine expression pattern at this stage. *Shh* was shown to be expressed within the posterior of the limb bud as well as the brain and the gut (Figure 5.7 B [v-vii]) similar to what is reported within the literature (Echelard et al. 1993). *In situ* hybridisation of embryos with probes designed against both *Lmbr1* and *Rnf32*

resulted in embryos showing staining across the embryo suggesting these genes could be ubiquitously expressed (Figure 5.7B [ii-iii]). RT-PCR (for primers see table 3.1 'RT-PCR analysis') was also employed using RNA generated from tissue of the limb, the head and the rest of the embryo; both genes were shown to be expressed within all these tissues. Upstream of *Lmbr1* two genes were examined, *Nom1* and *Mnx1*. *Nom1* was again found to result in an *in situ* and RT-PCR pattern suggesting ubiquitous expression (Figure 5.7B [i]). *Mnx1* was found to have some expression within the posterior of the limb bud (Figure 5.7B [iv]) a pattern which has previously been reported by (Rock et al. 2007). In addition, (Thaler et al. 1999) suggested that *Mnx1* is also expressed within motor neurons.

The expression pattern of three genes downstream of *Shh* was also determined by performing *in situ* hybridisation on E11.5 embryos. In accord with published data (Paek et al. 2012) both *Cnpy1* and *En2* were found to be expressed within the mid-brain hind-brain junction suggesting the expression pattern seen within the SBLac3-24 and SBLac1-66 tetraploids could reflect an enhancer element controlling these two genes (Figure 5.7B [ix-x]). *in situ* and RT-PCR suggested that *Rbm33* was again potentially ubiquitously expressed throughout the embryo (Figure 5.7B [viii]). The ENCODE project has also enabled examination of the genomic region surrounding *Shh* in detail (Consortium 2011; Shen et al. 2012). RNA polII sites which demarcate actively transcribed genes were examined in ES cells and shown to be present at the promoters of *Nom1*, *Lmbr1*, *Rnf32* and *Rbm33* suggesting these all indeed have active promoters (Figure 5.8).

These expression patterns served to show that the surrounding genes neither reflected *Shh* expression patterns nor show a general up-regulation in *Shh* domains, suggesting that apart from some potential ZRS activation of *Mnx1* the surrounding genes are shielded from the activity of the *Shh* enhancers.

5.9.2 Is the *Zrs* capable of acting on *Mnx1*?

In order to determine whether the expression of *Mnx1* found within the limb was driven by the *Zrs*, *in situ* hybridisation was also performed on embryos from a line where Shh is misexpressed within the limb as a result of modification to the *Zrs*

(See section 3.22). Mnx1 was similarly misexpressed suggesting that the *Zrs* does indeed function beyond its regulatory domain and is also capable of functioning on another developmental gene (Figure 5.7B [iv]).

Further analysis was used to examine tissue from either the anterior or posterior of the limb bud. RT-PCR showed that both *Shh* and *Mnx1* were expressed only within the posterior of the limb bud. A *β-actin* control was shown to be expressed equally within both regions (Figure 5.7C). This adds more evidence to *Mnx1* expression being driven by the ZRS as it is expressed within the same region at the same time-point.

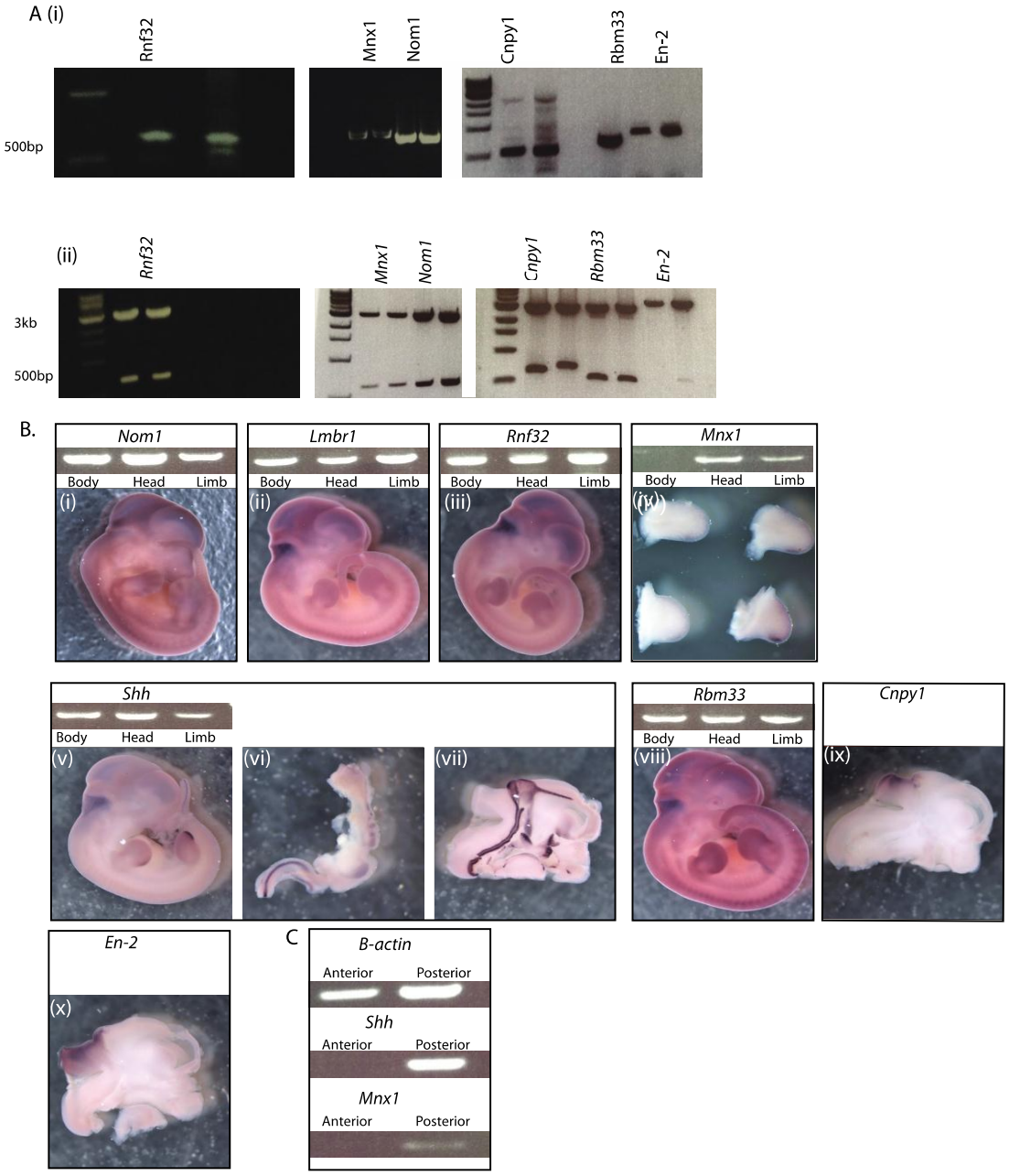


Figure 5.7: Expression of genes neighbouring the *Shh* regulatory landscape

A. (i) 500bp PCR fragments were generated for the genes *Rnf32, Mnx1, Nom1, Cnpy1, Rbm33* and *En2* (labelled).
(ii) PCR fragments were cloned into the p.Bluescript vector (3 kb) and a restriction digest with *Not1* and *Sal1* used to excise the 500 bp fragment.

B. *In situ* hybridisation of 11.5 day embryos for the genes *Nom1, Lmbr1, Rnf32, Shh* and *Rbm33*. *Mnx1* staining performed on limb buds from wild-type (right) and *Zrs* mutant (left). *Shh* staining on guts and sagittal head of 11.5 day embryo also shown. Staining of sagittal head of 11.5 days embryo rather than whole embryo is shown for *Cnpy1 and En-2*. RT-PCR on cDNA generated from whole wild-type embryos minus head, head only and limb only for *Mnx1, Nom1, Lmbr1, Rnf32, Shh and Rbm33* also shown above relevant *in situ* image.

C. RT-PCR on cDNA generated from tissue of the anterior and posterior of the limb bud for the genes *Shh* and *Mnx1* also shown with that of a positive control *Actb*.

5.10 Is there a topological domain surrounding *Shh*?

While this work was being undertaken, a topological domain stretching from *Shh* past *Lmbr1* has been reported for the *Shh* regulatory region (Figure 5.8A) with a definite boundary at the *Shh* end of the region and a less defined boundary within *Lmbr1* (Smallwood and Ren 2013). Data generated from the tetraploid complementation embryos with regards boundaries suggests that the topological domain suggested could indeed be correct. Topological boundary regions are believed to be enriched for the insulator binding protein CTCF, as well as housekeeping genes, transfer RNAs and short interspersed element (SINE) retrotransposons (Dixon et al. 2012) which prevent illegitimate enhancer-promoter interactions**.** Many known insulator or barrier elements are bound by the zinc-finger containing protein CTCF. Within the *Shh* and *Lmbr1* region a number of such sites were found, mainly at either end of the regulatory domain. No direct relationship was identified between the position of the CTCF sites and the loss of enhancer activity (ENCODE 2011). For example, there is a CTCF site between SBLac2-200 and SBLac2-21 (Figure 5.8) but *Shh* expression pattern does not appreciably change (Figure 5.3). At the *Shh* end of the region, a CTCF site is found between *Shh* and SBLac3-41 (Figure 5.8B) which could play a role in the down regulation of expression exhibited by this insertion (Figure 5.4). The gradual loss of expression toward the limits of the *Shh* topological corresponds with multiple CTCF sites; however there is no direct correlation suggesting there are additional mechanisms that limit the activity of enhancers within a regulatory domain (Figure 5.8C).

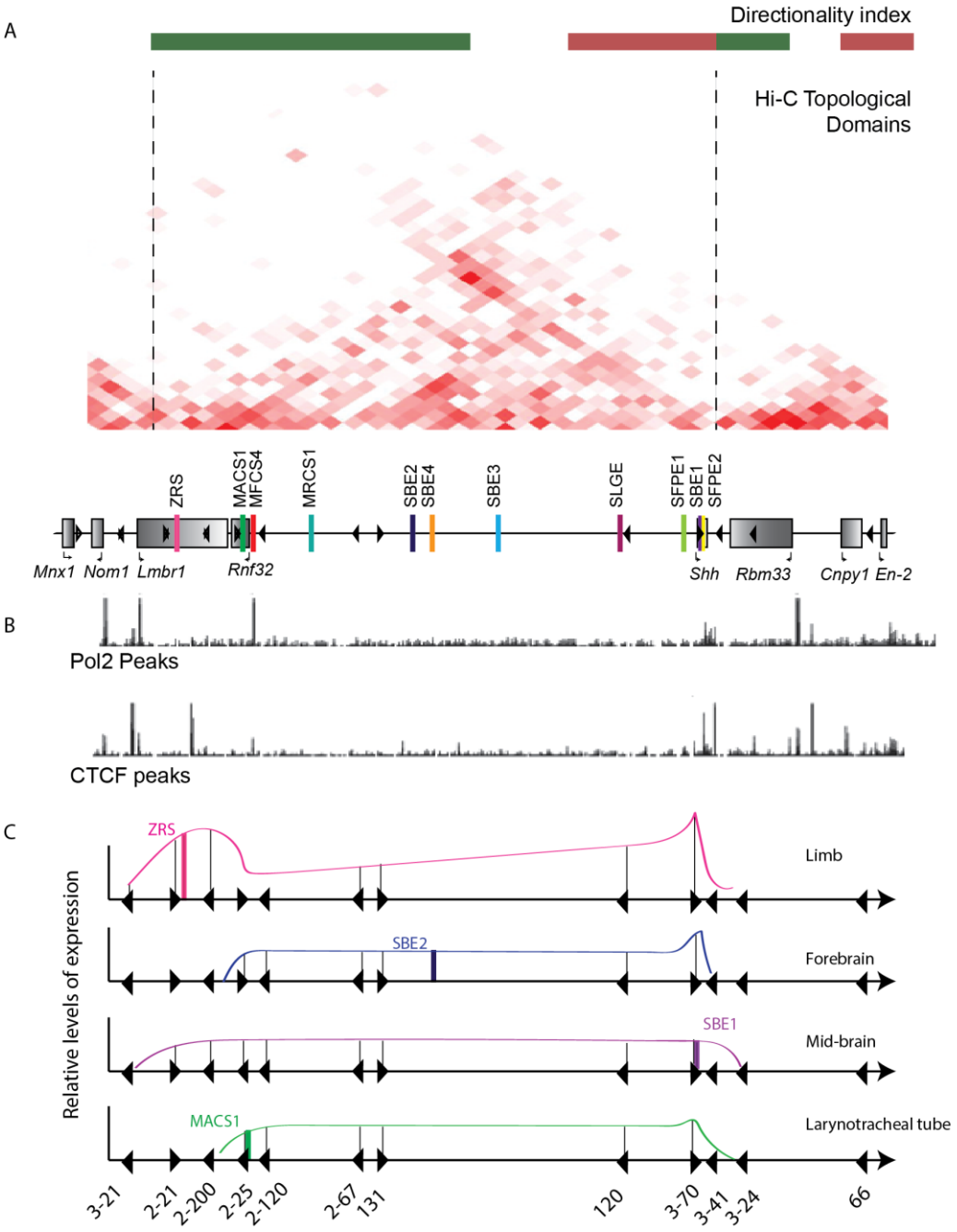


Figure 5.8: Summary of the topological domains surrounding the Shh regulatory region

A. Hi-C analysis from mouse ES cells (taken from the Mouse ENCODE website), and the boundaries of the topological domain determined by the directionality index and marked by dotted lines. The relationship between the positions of these features and the genes and insertions within the *Nom1* to *En2* region are indicted.
B. Below the genes is the track showing the positions of RNA polymerase II and CTCF sites (marked with black lines) in E14.5 limb buds taken from ENCODE.

C. A summary of relative expression activity driven by individual enhancers in particular tissues (*Zrs* driving expression in the limb bud, *SBE2* in the forebrain, *SBE1* in the midbrain and *MACS1* in the laryngotracheal tube) at each of the SBLac insertions. The solid coloured lines show an estimate of the relative expression levels throughout the genomic interval. Expression at the 5′ end shows that reduction inexpression occurs over an interval of greater than 100 kb, whereas limb bud expression is still detectable at the furthest insertion site.

5.11 Conclusions

By generating a series of tetraploid complementation embryos from cells carrying insertions spanning the *Shh* regulatory locus it has been possible to construct a map of reporter gene activity. Many of the regulators responsible for *Shh* embryonic expression within tissues such as the CNS, gut and limb are well characterised (Epstein, McMahon, and Joyner 1999; Lettice et al. 2002; Sagai et al. 2004; Sagai et al. 2005; Jeong et al. 2006; Sagai et al. 2009) and the combination of these can account for the expression pattern seen within this data.

Reporter gene activity across the gene desert suggests that position plays a minor role in enhancer activity. Regardless of where the reporter gene was positioned across the desert it was shown that there was very little change in expression pattern. This suggests that the entire region is open for activation. One potential issue with this is that some enhancers in particular SBE1 have been reported to function alongside redundant sometimes unidentified enhancers, however as only SBE1 functions at the E11.5 time point it is expected this should not influence the reporter output. These data contrast completely with that which suggests that cis-regulatory landscapes function as ‘holoenhancers’ or archipelagos (Montavon et al. 2011; Marinic et al. 2013). In these two models it has been suggested that multiple regulatory elements, many of which lie inside genes, act together to generate a complex developmental expression pattern. The enhancer activity and the expression pattern are highly dependent on the position of the gene relative to the regulatory elements suggesting a complex interaction of the regulatory elements. The *Shh* developmental locus therefore may represent a more common genomic composition (Symmons and Spitz 2013) relying on a regulatory domain primarily composed of regulators that contribute to the spatiotemporal expression pattern as a summation of the individual enhancer activities.

Variation is seen in terms of the distance in which the various *Shh* enhancers work (Figure 5.8) Both the *Zrs* and *SBE1* were shown to be capable of driving expression at large distances even though in the case of *SBE1* it is in fact located close to the promoter of the gene its acting upon. This suggests that long-range activity is something inbuilt into enhancers. The *MACS1* enhancer which drives expression within the gut was shown to act at long-range and in fact does this better than at short range. *SBE2*, *SBE3* and *SBE4* which are located within the middle of the region have been shown to be the least robust and do not appear to function as well at long distances.

The *Rnf32* gene which is located within the middle of the *Shh* landscape does not appear to be affected by the enhancers surrounding it and positioned within it. It has been shown that the actual body of the gene is open for enhancer activity suggesting there is something different about the promoter which shields it from activity. This is a phenomenon that may be quite common within developmental landscapes. Both upstream and downstream of the *Shh* locus lie several other genes; the *Shh* enhancers do not appear to function much downstream of *Shh* suggesting the presence of a fairly strong genomic boundary. Although some activity is detected 5 kb downstream by the time the next gene, *Rbm33* is reached 100kb further downstream it has been lost entirely and instead an entirely different regulatory landscape identified (Figure 5.4). The *Rbm33* gene itself appears to be in range of a putative regulatory element driving expression of *Cnpy1* and *En2* while like *Rnf32* its promoter remains refractory to its influence. This suggests that a scenario whereby a gene's promoter being intensive to the enhancers within or close by it (as shown with both *Rnf32* and *Rbm33*) could be a common occurrence within regulatory landscapes.

The boundaries surrounding the *Shh* regulatory landscape do not appear to be absolute. At the *Lmbr1* end of the domain enhancer function is found to tail off incrementally within the gene (Figure 5.3). Laryngotracheal expression is also reduced between SBLac2-25 and SBLac2-200, while the expression in the lung primordium is lost between SBLac 2-200 and SBLac2-21. Finally, the ventral midbrain and the limb expression levels are drastically reduced between SBLac2-21

and SBLac3-21 although low levels of *Shh* in the limb are still detectable within this insert and potential the next SBLac3-119. The *Zrs* therefore appears to function upstream of the region where it is responsible for driving expression of *Mnx1* within the posterior of the hind limb. The *Zrs* appears to act on *Mnx1*, whether this for functional purposes however remains unclear. Patients with Currarino syndrome have been found to have mutations involving deletions within the *Mnx1* gene. These patients exhibit anorectal, sacral and pre-sacral anomalies however it has not been noted whether they show any signs of limb malformation (Holm et al. 2013). The topological domain previously suggested for *Shh* (Smallwood and Ren 2013) fits with the data generated with regard to boundaries (by the tetraploid embryos) (Figure 5.8).

6. Analysis of LHED mouse lines

6.1 Introduction

As well as allowing mapping of regulatory activity across a region, the presence of *LoxP* sites both within the initial LHED insertion vector and also within the transposed transposable element means the system also provides an experimental platform for to generate deletions in mice to look at loss of function effects (Kokubu et al. 2009) (Figure 4.1). This technology therefore allows regions within a locus to be removed to see how these deletions affect phenotype, as a result of the elimination of enhancer or insulator elements. By creating these deletions the dissection of enhancer function can be performed. Even as of yet unidentified elements can also be identified by this technology, as well as allowing the locus to be disrupted to see how this affects the rest of the region. Additionally, if inserted within the correct orientation such a deletion does not remove the *LacZ* reporter gene meaning that reporter gene expression can be used to monitor loss of function effects within a regulatory region.

As previously discussed (Section 2.8) a number of known enhancers reside within the region between *Shh* and the ZRS. Three insertions distributed across the *Zrs* to *Shh* region were therefore chosen to generate mouse lines. These were generated from the original Shh2_LHED insertion 3' to the *Zrs* and located approximately 700 kb (SBlac2-120), 500 kb (SBlac2-67) and 100 kb (SBlac120) from *Shh*. These three insertions when treated with *Cre* recombinase would allow nested deletions of firstly *MACS1* and *MFCS4*, and then all three epithelial enhancers and finally all three epithelial enhancers as well as the three forebrain enhancers. In the final deletion only the known enhancers located closest to *Shh* which drive expression within the midbrain and floor plate would remain, as well as the Zrs (Figure 6.1). In addition to removing the enhancers within the region these deletions would also serve to disrupt any requirement for the large distance found between *Shh*

and the Zrs and potentially reveal the location of as yet unidentified enhancers or regulators.

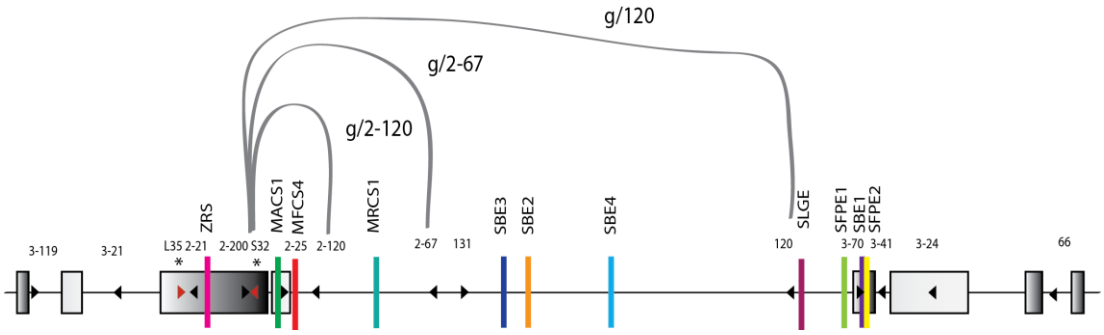


Figure 6.1: Deletions within the *Shh* region

Schematic diagram representing the location of the three deletions generated; g/2-120, g/2-67 and g/120 within the *Shh* region. Enhancers shown as coloured rectangles with the various SBLac insertions shown as triangles and the original insertion sites as red triangles demarcated with *.

6.2. Mouse line production

ES cells were injected into C57Bl/6 embryos to generate chimeras. Chimeras were then crossed to C57Bl/6 mice in order to produce F1 offspring, which was confirmed by PCR using primers against the *LacZ* reporter gene and designated with SBLac and the insertion number e.g. SBLac120. F1 mice from each of the SBLac lines were subsequently crossed to germ line Cre animals (g/Cre) in order to induce a deletion between the two *LoxP* sites (See materials for further details). Primers (Table 3.1 'Genotyping LHED deletion 3'') were designed either side of the deleted region, which would produce a positive PCR band only if the intermediate region was removed and tail tip DNA was used to genotype mice. Mice positive for the deletion are designated; SBLac$^{\Delta chr5:\ start\ of\ deletion\text{-}end\ of\ deletion}$ e.g. SBLac120$^{\Delta chr5:\ 29567195\text{-}\ 28793641}$ however for simplicity will be referred to as g/ and their SBLac insert number e.g. g/120 for the remainder of this text.

6.2.1 Expression analysis

F1 mice from each of the insertions before and after *Cre recombinase* mediated deletion were crossed to CD1 females in order to generate embryos heterozygous for

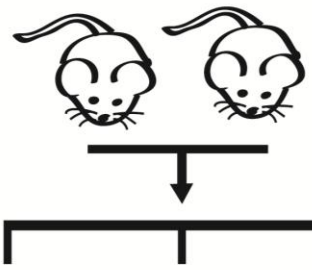
the *LacZ* reporter gene. These embryos were then harvested at stages between E9.5 and E14.5 and stained using X-gal in order to map how reporter gene expression and thus the activity of the *Shh* enhancers varied across time and also to determine the effect of the three deletions on the expression pattern.

6.2.2 Phenotype analysis

Heterozygous F1 mice from each of the deletion lines were found to be to be viable and developmentally normal. This was not surprising considering the *Shh* null mutant is also viable and developmentally normal as a heterozygote (Chiang et al. 1996). Therefore matings were set up to produce homozygotes for each line. However genotyping from a number of litters from each of the lines suggested that no homozygous pups were found shortly after birth or at weaning it was therefore hypothesised that the phenotype could be lethal at embryonic or neonatal stages (See Figure 6.2 for numbers). Additional matings were therefore set up with the intent of harvesting the pups prior to birth at either early stages to look at LacZ expression or later stages of E15.5 and E17.5 in order to examine phenotype. Of the three lines, homozygotes were only identified at E12.5 for the largest deletion g/120, this was as a result of breeding issues rather than embryo death. A large number of reabsorption sites were however identified for all three lines at later stages of development suggesting embryo death. All the deletions encompass the *Rnf32*, which has so far not been characterised, however is expressed ubiquitously during development. It is therefore possible that homozygous deletion of *Rnf32* could be responsible for the embryonic lethal phenomenon exhibited in all three of the lines. In order to combat this, heterozygous male mice from each of the deletion lines were crossed with female mice carrying a *Shh* null chromosome, in order to generate compound heterozygous mice with one remaining copy of *Rnf32* and only capable of *Shh* expression from the enhancers not located within the deleted region, these were designated with $Shh^{-/+}$ to indicate the presence of a *Shh* null allele thus a mouse with the genotype $g/120^{+/-}$; $Shh^{-/+}$ reflects a compound heterozygote for the g/120 line (with + being the wild type chromosome in both cases). Genotyping was undertaken on a number of litters post-birth (Figure 6.2) for each of the lines and revealed no

mice which were positive as compound heterozygotes. This again suggesting a lethal phenotype was occurring. Embryos were therefore again harvested at early stages (E11.5/E12.5) to look at LacZ expression and also later embryonic stages of E17.5 to examine phenotype. Compound heterozygotes for each of the lines crossed with the *Shh* null were identified using this method (Figure 6.2).

g/2-120+/- x g/2-120+/-

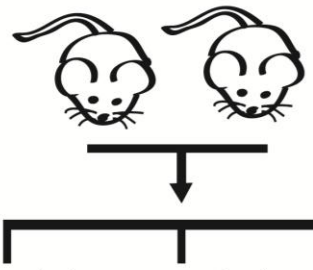


g/2-120+/+ (wt) g/2-120+/- (het) g/2-120-/- (hom)

Observed/expected

Birth (N=10)	3/2.5	7/5	0/2.5
E17.5 (N=16)	5/4.75	11/8	0/4

g/2-67+/- x g/2-67+/-



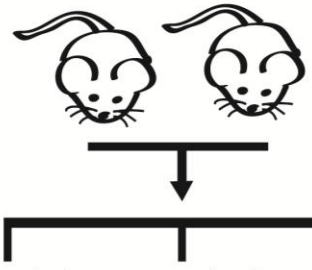
g/2-67+/+ (wt) g/2-67+/- (het) g/2-67-/- (hom)

Observed/expected

Birth (N=19)	10/4.75	9/9.5	0/4.75
E17.5 (N=13)	7/3.25	6/6.5	0/3.25

* used to denote varied craniofacial defect** used to denote gut phenotype and *** for pharyngeal/facial defect

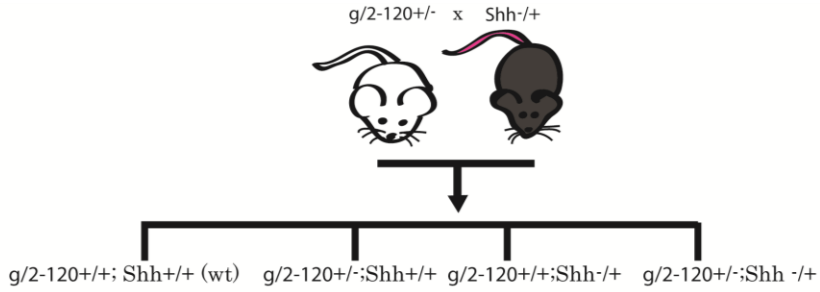
g/120+/- x g/120+/-



g/120+/+ (wt) g/120+/- (het) g/120-/- (hom)

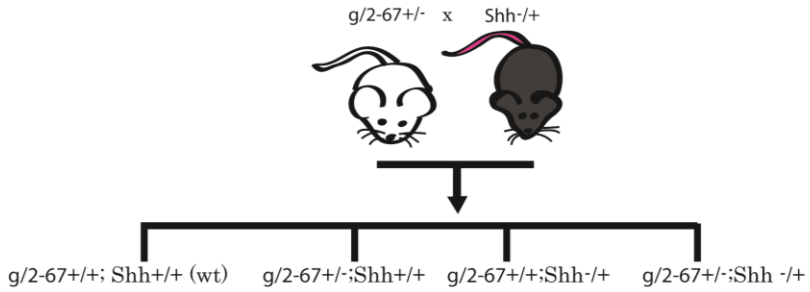
Observed/expected

Birth (N=17)	6/4.25	11/8.5	0/4.25
E17.5 (N=11)	7/2.75	4/5.5	0/2.75
E12.5 (N=10)	2/2.5	3/5	5/2.5*



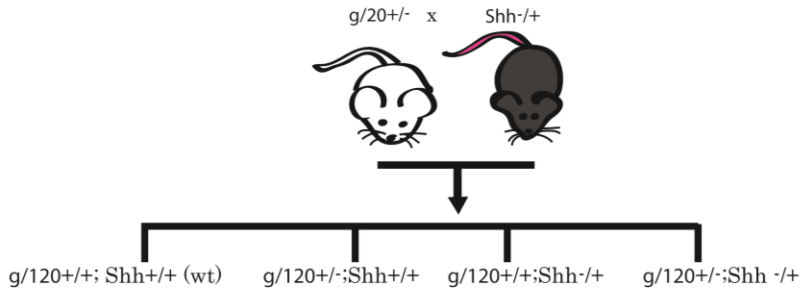
Observed/expected

Birth (N=25)	9/6.25	12/6.25	4/6.25	0/6.25
E17.5 (N=24)	6/6	8/6	7/6	3/6**



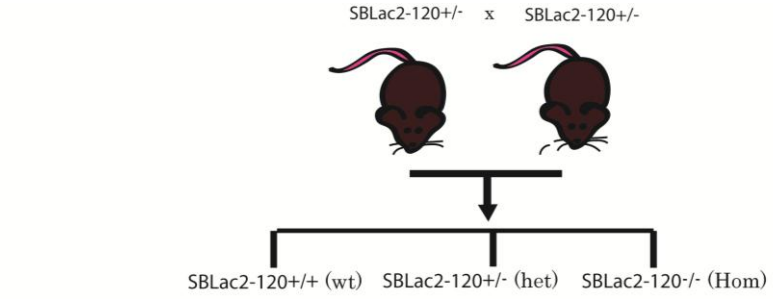
Observed/expected

Birth (N=12)	4/3	4/3	4/3	0/4
E17.5 (N=23)	8/5.75	6/5.75	3/5.75	3/5.75***
E12.5 (N=7)	0/1.75	2/1.75	4/1.75	1/1.75***

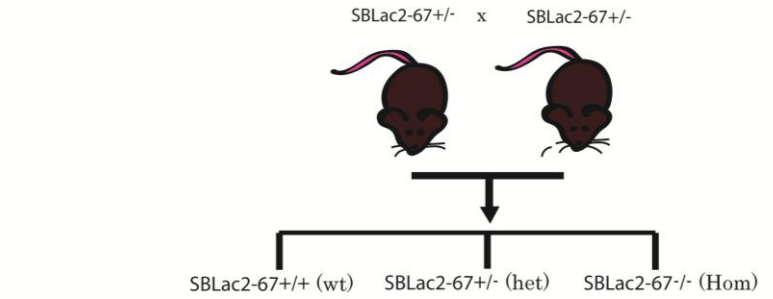


Observed/expected

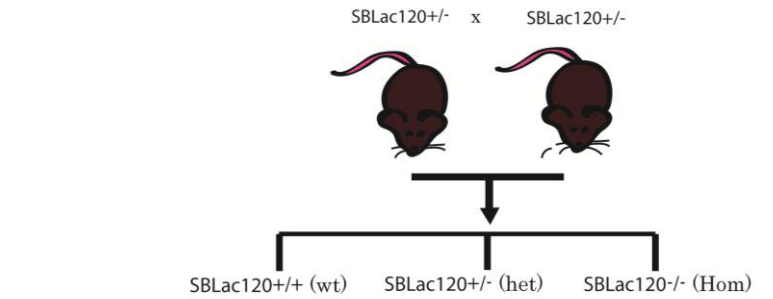
Birth (N=7)	3/1.75	2/1.75	2/1.75	0/1.75
E17.5 (N=24)	5/6	7/6	8/6	4/6*
E12.5 (N=6)	1/1.725	0/1.25	1/1.25	4/1.25*



Observed/expected			
E11.5-E17.5 (N=16)	7/4	7/2	2/4



Observed/expected			
E11.5-E17.5 (N=22)	9/5.5	10/11	3/5.5



Observed/expected			
Birth (N=14)	5/3.5	9/7	0/3.5
E11.5-E17.5 (N=18)	6/4.5	6/9	6/4.5*

Figure 6.2: Mouse crosses and resulting pups

Mouse crosses set up for each of the lines and actual pup numbers shown for the various stages as well as calculated expected numbers based on observed total pup numbers. Lack of expected homozygotes/compound heterozygotes suggests embryo death. * used to denote varied craniofacial defect** used to denote gut phenotype and *** for pharyngeal/facial defect

6.3 How do deletions within the *Shh* region affect enhancer activity and development?

6.3.1 *LacZ* reporter gene analysis of SBLac lines

In order to provide a reference to compare each of the lines to embryos from the SBLac lines carrying a wild-type chromosome (prior to deletion) were examined. The staining pattern within these embryos was found to be consistent across the three insertions. All three insertions exhibited a similar expression pattern with regards to the tissues of expression with only minor variation in the levels of expression, similar to what was previously found within the tetraploid complementation embryos (See previous chapter). The SBLac2-67 line, is therefore, shown as a representative for all three (Figure 6.2). Within the 9.5 day embryos, expression was found to occur within the brain and also within tissues of the floor plate and notochord as well as within the gut. By E10.5, LacZ expression was found extensively throughout the central nervous system, the gut and also within the posterior of the limb bud. At E11.5,a similar expression pattern was found, with staining present within the CNS, the gut and the limb; it was also possible to note expression within the pharyngeal arches. The guts were also dissected from the E11.5 day animals with strong expression found throughout the structures of the lung, stomach and oesophagus. By E12.5, expression began to be detected within structures of the head, it now being possible to detect staining within the whisker buds, limb expression was also found to start to fade. By E13.5 the expression within the limb was severely reduced and by E14.5 it had disappeared completely. Strong expression was also found within the umbilicus at E13.5. At E14.5, LacZ expression was found within individual hair follicles on the skin, it was also found strongly within the stomach and covering the taste buds of the tongue.

One limitation of using *LacZ* to monitor enhancer activity is perdurance of β-galactosidase activity may give an erroneous impression of a longer duration of activity. It is therefore unknown whether the expression pattern found within the embryos reflects the actual presence of RNA transcripts, however as expression is shown to turn on and off and also as expression for example in the limb is shown

only within the lineage of cells also known to express SHH this does not appear to be a problem. Employing RT-PCR or *in situ* hybridisation would help to confirm this.

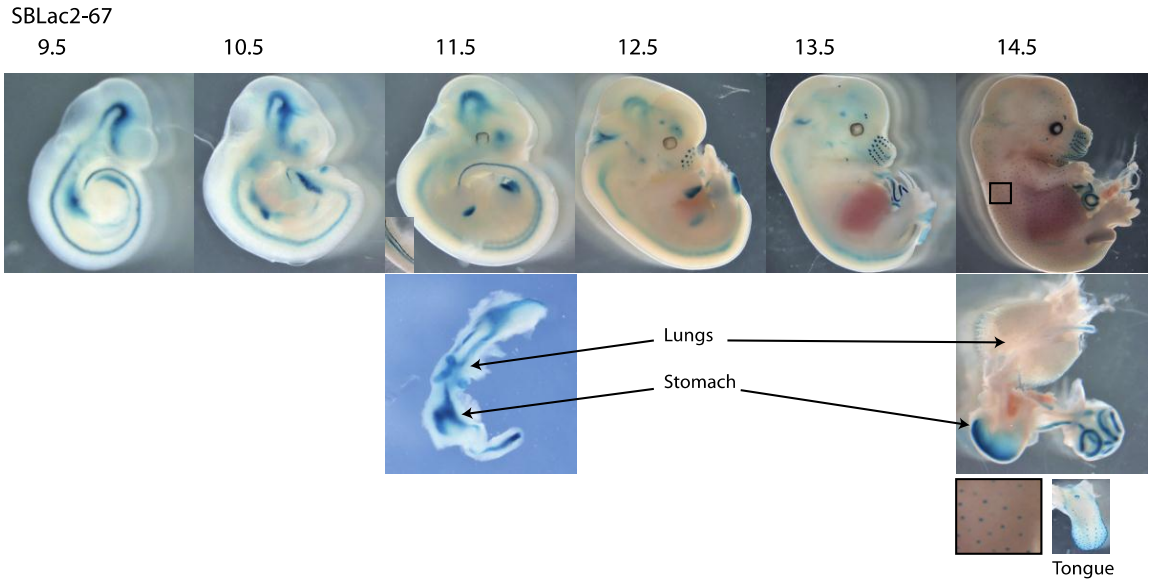


Figure 6.3: Reporter gene activity in a LHED control line

LacZ expression analysis of SBLac2-67 heterozygous LHED carrying embryos prior to *Cre* deletion, from stages E9.5 to E14.5. Notochord/floor plate and dissected gut is also shown for E11.5 days and gut and tongue for E14.5 days. Magnification of skin of E14.5 day embryo shown in box in bottom right hand corner with small box on embryo reflecting where image is taken from.

6.3.2 LacZ reporter gene analysis of g/2-120

The mouse line g/2-120 was generated which contains a deletion of approximately 80 kb of DNA encompassing the 3' end of the *Lmbr1* gene, *Rnf32* and two known *Shh* enhancers *MACS1* and *MFSC4* which are believed to control *Shh* expression within the epithelium of the gut and pharynx. This was confirmed by PCR on tail tip DNA, with primers designed either side of the deleted region.

Heterozygous mice were harvested between the stages of E9.5-E14.5 and analysed for LacZ expression. Embryos were found to recapitulate the expression pattern found within the control SBLac mice in all the tissues of the CNS and the limb throughout the stages (Figure 6.4).

The major difference found was a reduction in expression within the gut of the embryo of the g/2-120 line. This deletion was found to result in the loss of the majority of LacZ expression within the lungs and oesophagus leaving only low-level expression relative to other tissues. Expression was also found to be lost from the pharyngeal arches, however it was possible to show by dissecting the heads of these animals that, like the gut, residual expression remained (Figure 6.5).

Later stage embryos containing the g/2-120 deletion were also examined to determine expression pattern. E12.5 embryos were found to have a similar expression pattern within the face to those carrying a wild-type (undeleted) chromosome, with expression found within the whisker buds. At both E13.5 and E14.5, embryos also appeared to be fairly similar to the control mouse line and it was evident that reporter gene expression was still present within the developing follicles of the skin. At E14.5, embryos were also dissected to remove the guts with a similar tissue distribution identified to the E11.5 guts whereby a global reduction in expression was found as compared to the control. The tongue was also examined at this stage and shown to have undergone a loss of reporter expression from the posterior of the tongue although not the anterior thus suggesting a previously undescribed regionalization of expression.

MACS1 and *MFCS4* have been reported within the literature as driving expression within the gut and pharyngeal epithelium (Sagai et al. 2009). The reduction in expression within the tissues of the gut and pharyngeal structures within the g/2-120 line suggests this is indeed the case. However, the presence of remaining residual staining within the tissues of the gut and pharynx even after deletion of the enhancer responsible suggests that a shadow enhancer is present out with the deleted region which can drive expression within these tissues. A second gut enhancer SLGE has been identified close to *Shh* (Tsukiji, Amano, and Shiroishi 2013) and thus unaffected by the g/2-120 deletion, which is likely to be responsible for driving expression within the gut. No other enhancers have been reported to drive pharyngeal expression.

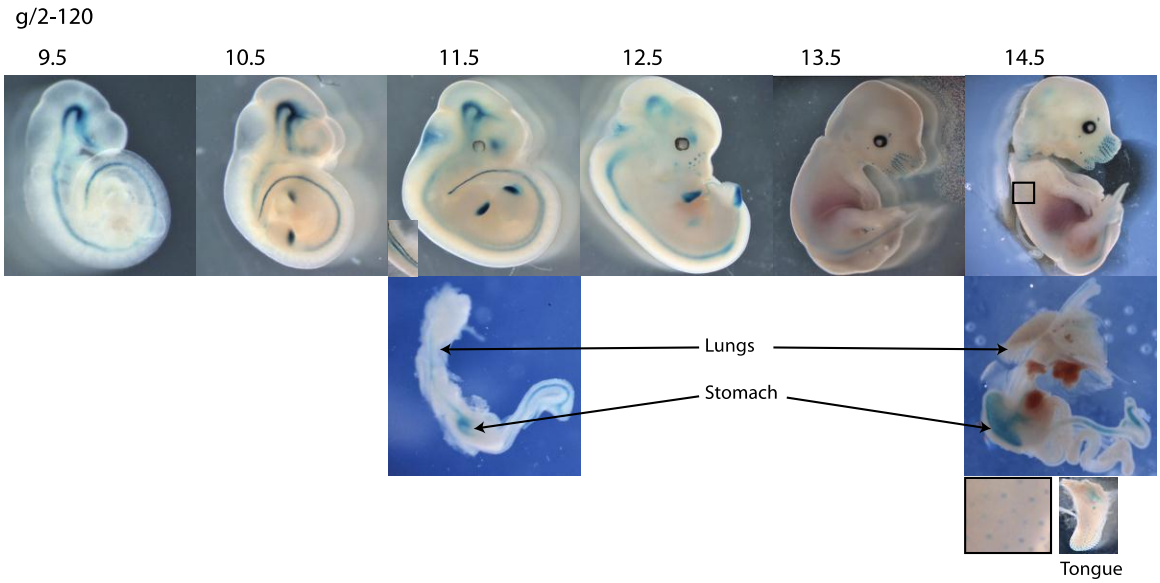


Figure 6.4: Reporter gene activity in the heterozygous g/2120 mouse line

LacZ expression analysis of following g/2-120 deletion. Embryos stained from E9.5 to E14.5 including dissected guts and neural tube at E11.5 and gut and tongue at E14.5. Skin follicles also shown as above.

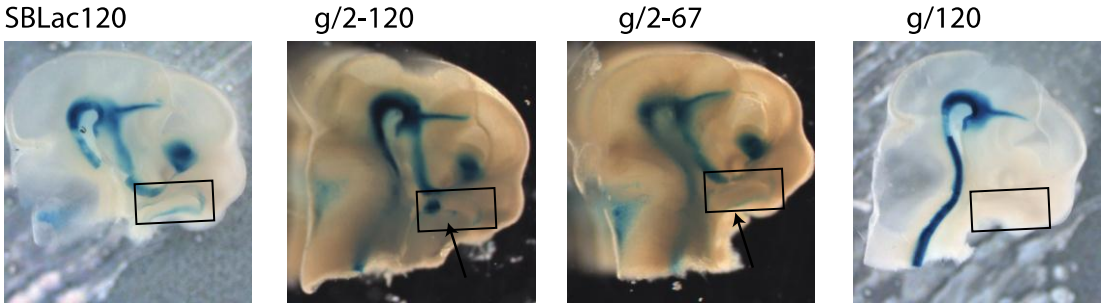


Figure 6.5: Reporter gene activity within sagittal head sections from LHED mouse lines

Sagittal section of embryo head from SBLac120, g/2-120, g/2-67 and g/120 lines. Expression shown within the forebrain, midbrain, hindbrain and pharyngeal structures of SBLac120. Region of pharyngeal expression highlighted by black box, it is shown to be present within the SBLac120 control line and although reduced within g/2-120 and g/2-67 some LacZ expression remains (black arrows). No pharyngeal expression is however found within the g/120 line.

6.3.3 Phenotype analysis of the g/2-120 line

Compound heterozygous mice (*g/2-120*$^{+/-}$*; Shh* $^{-/+}$) were identified at E17.5 from crosses of the g/2-120 line with the *Shh* null line (Figure 6.2). Outwardly these animals appeared phenotypically normal with no defects noted within the structures of the face or limb (associated with *Shh* misexpression). As the g/2-120 deletion had

resulted in removal of the gut enhancer MACS1 and expression analysis suggested that loss of expression had occurred with those tissues, the guts from *g/2-120*$^{+/-}$*; Shh*$^{-/+}$ embryos were dissected. Phenotypic differences were observed within the guts of the g/2-120$^{+/-}$; Shh $^{-/+}$ animals. The stomachs of the animals were found to be much smaller than litter mate controls and were also abnormally shaped. The intestines of these animals were measured and found to be approximately 20% shorter than those of the controls (Figure 6.6A).

The guts, stomachs and lungs taken from E17.5 *g/2-120*$^{+/-}$*; Shh*$^{-/+}$ animals were fixed and embedded before being sectioned. H&E staining was undertaken on slides in order to look at gross morphology of the tissues. There were some obvious morphological differences with respect to the mutants and controls. The stomachs of mutant animals were found to have an altered morphology as compared to the wild-type with the epithelium found to be enlarged and overgrown (Figure 6.6B [ii] 6.6C [ii]). Within both the large and small intestine of the mutant eosin staining was seen between villi, which does not occur within the control animal (Figure 6.6B [i] 6.6C [i]). Eosin is known to stain basic structures such as the cytoplasm, cell wall and extracellular fibres thus suggesting intestine of the *g/2-120*$^{+/-}$*; Shh*$^{-/+}$ animals is becoming overgrown. Lungs of both mutant and wild-type mice were found to be fairly similar with normal morphology (Figure 6.6B [iii] 6.6C [iii]).

Shh null animals have been reported to display overgrown stomach epithelium, occlusion of the duodenum by overgrown villi and also intestinal transformation of the stomach (See section 2.4). To determine if the latter was also true within g/2-120, mutants alkaline phosphatase staining was undertaken. Alkaline phosphatase is normally present within the intestine but not the stomach. *g/2-120*$^{+/-}$*; Shh*$^{-/+}$ mutants and wild-type mice were both found to have high alkaline phosphatase activity within the intestine (Figure 6.6D [i] E [i]). Additionally, the over grown tissue within g/2-120$^{+/-}$; Shh$^{-/+}$ embryos was found to stain strongly for alkaline phosphatase. The stomach of the mutant were not found to have increased alkaline phosphatase staining as compared to a wild-type control (Figure 6.6D [ii] E[ii]) thus suggesting that in contrast to the *Shh* mutant, the *g/2-120*$^{+/-}$*; Shh*$^{-/+}$ mutant stomach is not a result of intestinal transformation. Antibody staining was also undertaken using an αSMA antibody in order to examine smooth muscle within the stomachs of both wild-type

(Figure 6.6F (i) and mutant animals as this is reported to be abnormal within the *Shh* null (Ramalho-Santos, Melton, and McMahon 2000), however again no differences were observed (Figure 6.6F(ii)).

Summary

The abnormal phenotype of the *g/2-120*$^{+/-}$*; Shh*$^{-/+}$ mice was found to resemble that of both *Wnt5a* knock-out and gain of function mice whereby a similar outgrowth defect within tissues including the gut is exhibited (Bakker et al. 2013). The phenotype observed within the *Wnt5a* study is hypothesised to be as a result of failure of the intestine to elongate suggesting a potential role in cell proliferation. Studies of targeted gene deletion and over-expression of *Wnt5a* using a lung-specific promoter showed that Wnt5a, Fgf10, and Shh signalling pathways are functionally interacting with *Wnt5a* negatively regulating Shh signalling in the lung (Li et al. 2005). It is therefore possible that the gut defect exhibited both within the *g/2-120*$^{+/-}$*; Shh*$^{-/+}$ mouse line, similarly to the *Wnt5a* mutant, could be as a result of changes to cell proliferation.

The gut and lung defects exhibited within the *g/2-120*$^{+/-}$*; Shh*$^{-/+}$ mutant are not as severe as those exhibited within the *Shh* null mutant. As previously shown by reporter gene expression however, the expression within the guts of g/2-120 heterozygotes was not completely lost, it could be possible this remaining expression, potentially from the SLGE enhancer is preventing a more severe phenotype. Within the *Shh* null mice, lungs fail to develop, within the g/2-120 heterozygous mice reporter gene expression also appears to be vastly down-regulated within the lungs suggesting enhancer function is also reduced. Surprisingly animals were found to have normal lung morphology. The lack of lung phenotype suggests firstly MACS1 may not be the only enhancer driving lung expression, and this could be the case as SLGE is also proposed to play a functional role. Additionally, this data suggests that SHH is still being expressed within the lungs of the g/2-120 mutant which is potentially enough to rescue any lung defect.

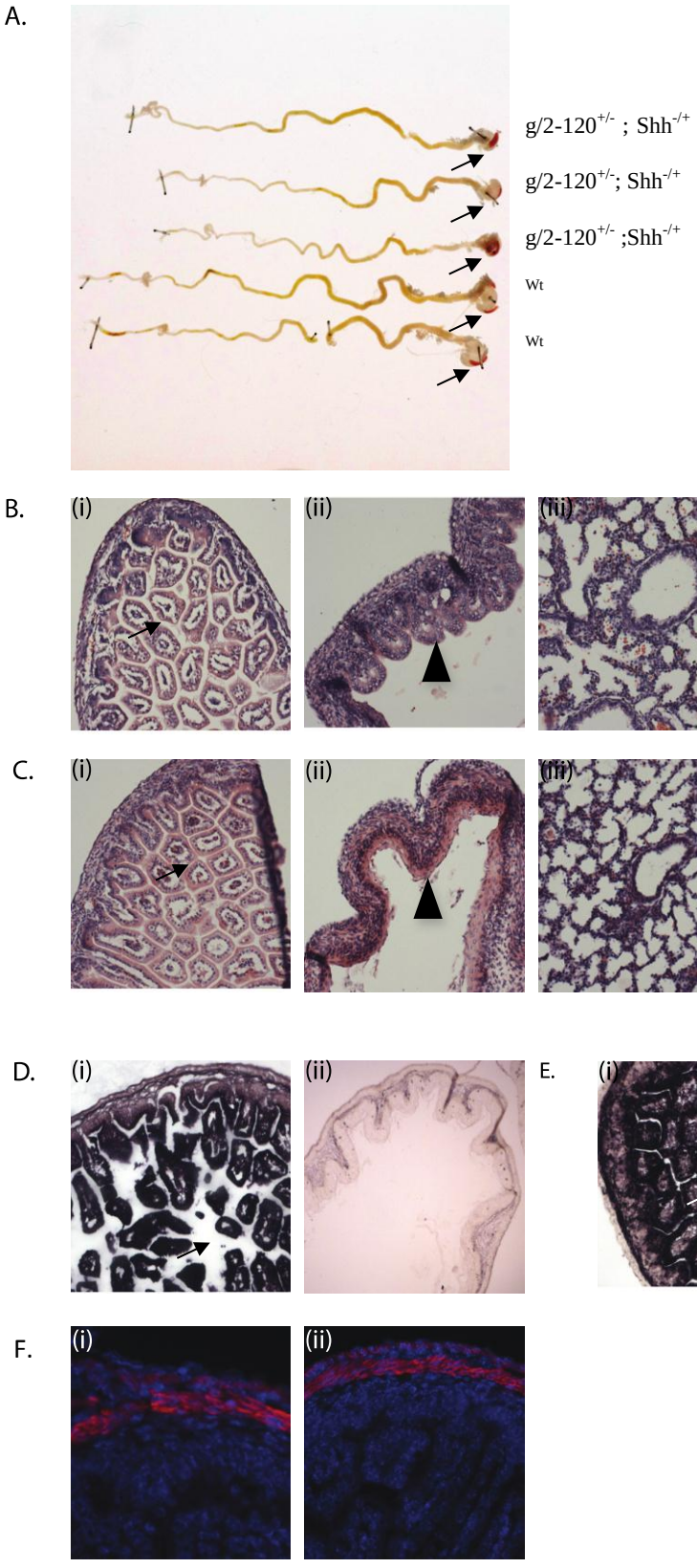


Figure 6.6: Phenotype analysis of the g/2-120 mouse line

A. Stomach and intestines of E17.5 mouse embryos in both *g/2-120*$^{+/-}$ *; Shh*$^{-/+}$ (top three) and wild-type (bottom two) mice. Noticeable size differences in the length of the intestine between mutant and wild type mice. Stomachs (black arrow) are also shown to be smaller and also misshapen in the mutant as compared to the wild type.
B. H&E staining of sections from E17.5 (i) wild-type intestine (ii) wild-type stomach lining (ii) wild-type lungs. Space is shown between the intestinal villi (black arrow), folding within the stomach is shown to be compact (Black arrow head). Lung morphology remains normal.
C. H&E staining of sections from E17.5 (i) *g/2-120*$^{+/-}$ *; Shh*$^{-/+}$ intestine (ii) *g/2-120*$^{+/-}$ *; Shh*$^{-/+}$ stomach lining and (iii) g/2-120$^{+/-}$; Shh$^{-/+}$ lung. Tissue is shown to be present within the villi of the intestine (black arrow), the stomach is shown to have lost the tight folding seen within the wild type mouse (Black arrow head). The lung morphology is shown to be normal.
D. Alkaline phosphatise staining of sections from E17.5 (i) wild-type intestine where staining is shown to be normal with space between villi (black arrow) (ii) stomach lining which is also shown to be normal.
E. Alkaline phosphatase staining of sections from E17.5 (i) *g/2-120*$^{+/-}$ *; Shh*$^{-/+}$ intestine which is shown to stain very darkly with the alkaline phosphatase with no space present between the villi (white arrow) (ii) *g/2-120*$^{+/-}$ *; Shh*$^{-/+}$stomach lining which is shown to stain comparably to the wild type control.
F. Antibody staining against smooth muscles within the stomachs of E17.5 (i) wild type stomach and (ii) *g/2-120*$^{+/-}$ *; Shh*$^{-/+}$ compound heterozygote. The stomachs of both animals are shown to be comparable.

6.3.4 LacZ reporter gene analysis of g/2-67 heterozygote

The g/2-67 line contains a deletion of approximately 248 kb of DNA including the 80 kb deleted within the g/2-120 mouse line as well as the oral epithelium enhancer *MRCS1* and a large amount of both repetitive and unique DNA of 'unknown function'. This was confirmed using the same PCR strategy on tail tip DNA as used with the g/2-120 line. Embryos were harvested at stages E9.5-E14.5 and stained for LacZ expression (Figure 6.7). Expression patterns of the g/2-67 embryos appeared to be very similar to those seen within the g/2-120 mouse lines.

All regions of known *Shh* expression were recapitulated apart from a reduction of pharyngeal expression and again a down-regulation of gut expression at E11.5 which was maintained until E14.5 (Figure 6.5/6.7). The major difference between g/2-120 and g/2-67 was the loss of the oral epithelium enhancer *MRCS1*. *MRCS1* is reported (Sagai et al. 2009) to be responsible for driving *Shh* expression within the epithelia of the incisor and molar teeth as well as the lingual papillae at E12.5. At E13.5, expression is within the whisker buds, tooth primordial and the rugae of the hard palate. Expression in these domains continues at least until the palatal structure is established at E15.5. At E14.5-15.5, the enhancer is also suggested to drive expression within the hair and nail buds (Sagai et al. 2009). It is evident that reporter gene expression is still present within the whisker buds of the g/2-67 E13.5 day

embryos and the developing skin follicles and tongue which had expression similar to that of the SBLac control line at E14.5 (Figure 6.7). The apparent maintenance of whisker bud expression within the g/2-67 embryos would suggest that *MRCS1* is either not required for driving expression within this tissue or alternatively is not the only enhancer driving this expression.

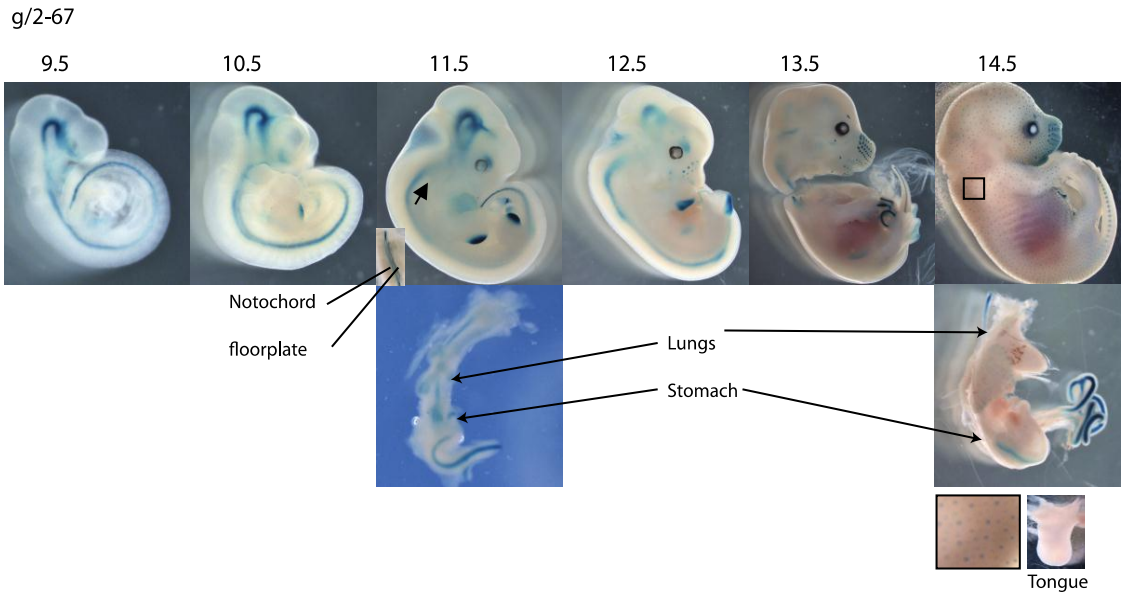


Figure 6.7: Reporter gene activity in the g/2-67 mouse line

LacZ expression analysis of following g/2-67 deletion. Embryos stained from E9.5 to E14.5 including dissected guts and neural tube at E11.5 and gut, skin and tongue at E14.5. As compared to the SBLac control; reporter gene expression is found to be reduced within the tissues of the gut and also lost from the pharyngeal arches (black arrow).

6.3.5 Phenotype analysis of the g/2-67 line

g/2-67$^{+/-}$ *; Shh*$^{-/+}$ compound heterozygous embryos were harvested at both E11.5 and E17.5 (Figure 6.2). E11.5 embryos were stained with X-gal and expression pattern examined. As compared to the g/2-67 heterozygote (Figure 6.8A[i], the compound heterozygote (Figure 6.8A[ii]) had normal expression pattern with regards to staining present within the CNS and limbs as well as reduced expression within the gut. However some differences were identified with regards to pharyngeal expression and in addition the development of the lower jaw appeared to be abnormal suggesting that the pharyngeal arches had not fused.

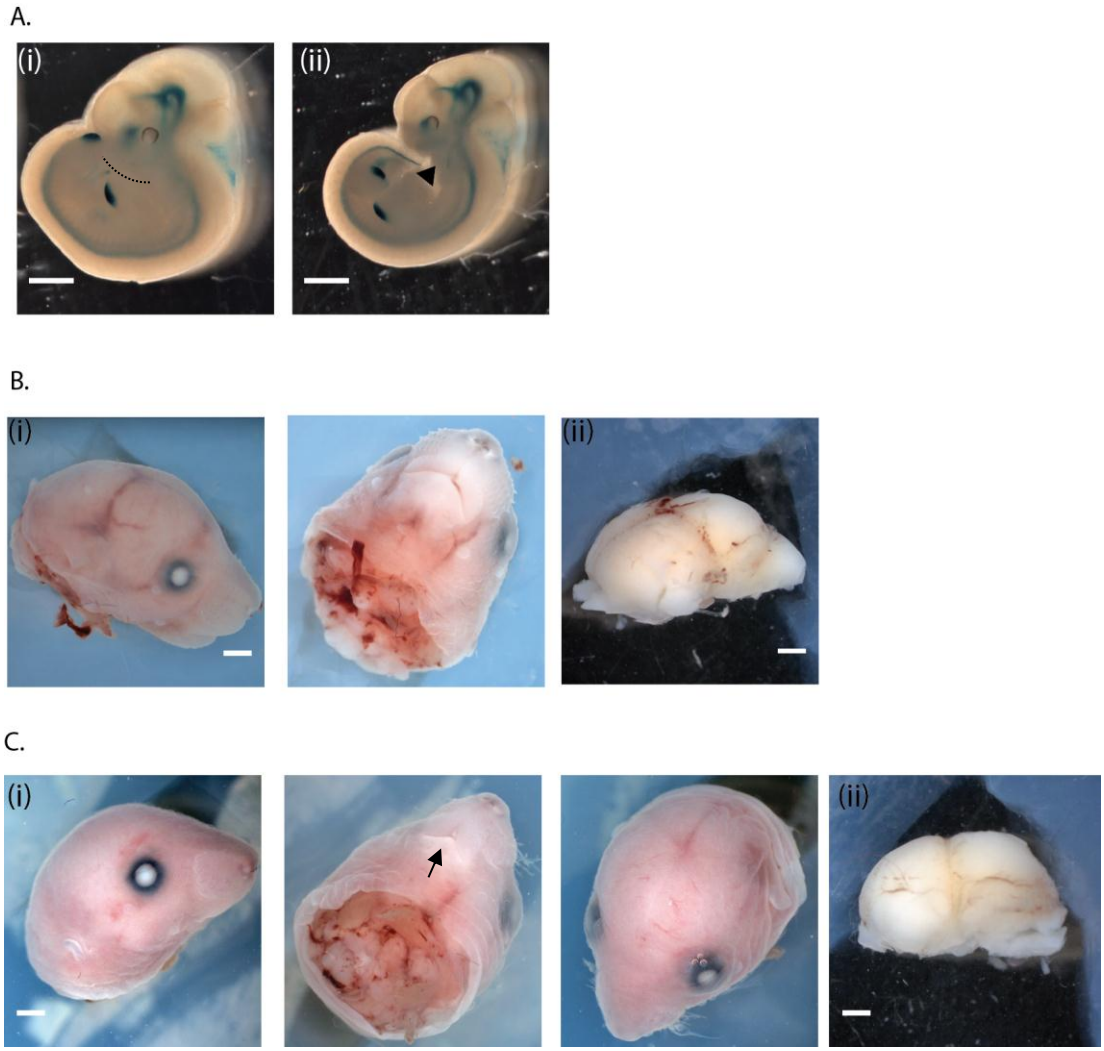


Figure 6.8: Phenotype analysis of the g/2-67 mouse line

A. X-gal staining of LacZ reporter gene expression within 11.5 (i) g/2-67 and (ii) *g/2-67*$^{+/-}$*; -Shh*$^{-/+}$ compound heterozygotes. Non-fused pharyngeal arch shown by black arrowhead, whereas the outline of normal jaw development in the heterozygous mouse is highlighted.
B. (i) Dissected head from E17.5 wild-type mouse embryo showing normal cranio-facial development. (ii) Dissected brain from E17.5 wild-type mouse embryo showing normal morphology.
C. (i) Dissected head from E17.5 g/2-67$^{+/-}$; Shh$^{-/+}$ mouse embryo with altered morphology including absent lower jaw (black arrow) and pointed face. (ii) Dissected brain from E17.5 *g/2-67*$^{+/-}$*; Shh*$^{-/+}$ mice exhibiting pointed face phenotype, morphology appears somewhat deformed however the brain structures are present.
Scale bars used to show 100 microns.

At E17.5 *g/2-67*$^{+/-}$*; Shh*$^{-/+}$ mutants were again found to exhibit a facial defect as compared to their wild-type litter mates (Figure 6.8B [i]). Within the *g/2-67*$^{+/-}$*; Shh*$^{-/+}$ embryos a phenotype was identified whereby the snout appeared pointed and the

lower jaw was underdeveloped, the tongue was also completely missing (6.8C [i[). The brain itself also appeared smaller and somewhat deformed in shape, however, it appeared to have all the correct structures present (Figure 6.9A) (Figure 6.8C[ii]) as compared to the brain of the wild type control (Figure 6.8C[iii]) (Personal communication Professor David Price). This could therefore be as a result of the shape of the skull influencing brain development. Brains from both wild type (Figure 6.9B) and *g/2-67*$^{+/-}$*; Shh*$^{-/+}$ animals (Figure 6.9C) were sectioned and stained with H&E. However, this failed to show any differences to brain morphology. The skull of wild type (Figure 6.9D) and g/2-67 compound heterozygote (Figure 6.9E) mice were also sectioned and stained, craniofacial malformations in the form of a shortened lower jaw and absent tongue were identified in the g/2-67 compound heterozygotes. Again the brain appeared to be abnormally shaped however all the structures were present. Gut defects were also identified within the *g/2-67*$^{+/-}$*; Shh*$^{-/+}$ mutant similar to that of the *g/2-120*$^{+/-}$*; Shh*$^{-/+}$ mutant whereby an increased amount of tissue was present within the intestinal walls (Figure 6.9 F[ii]).

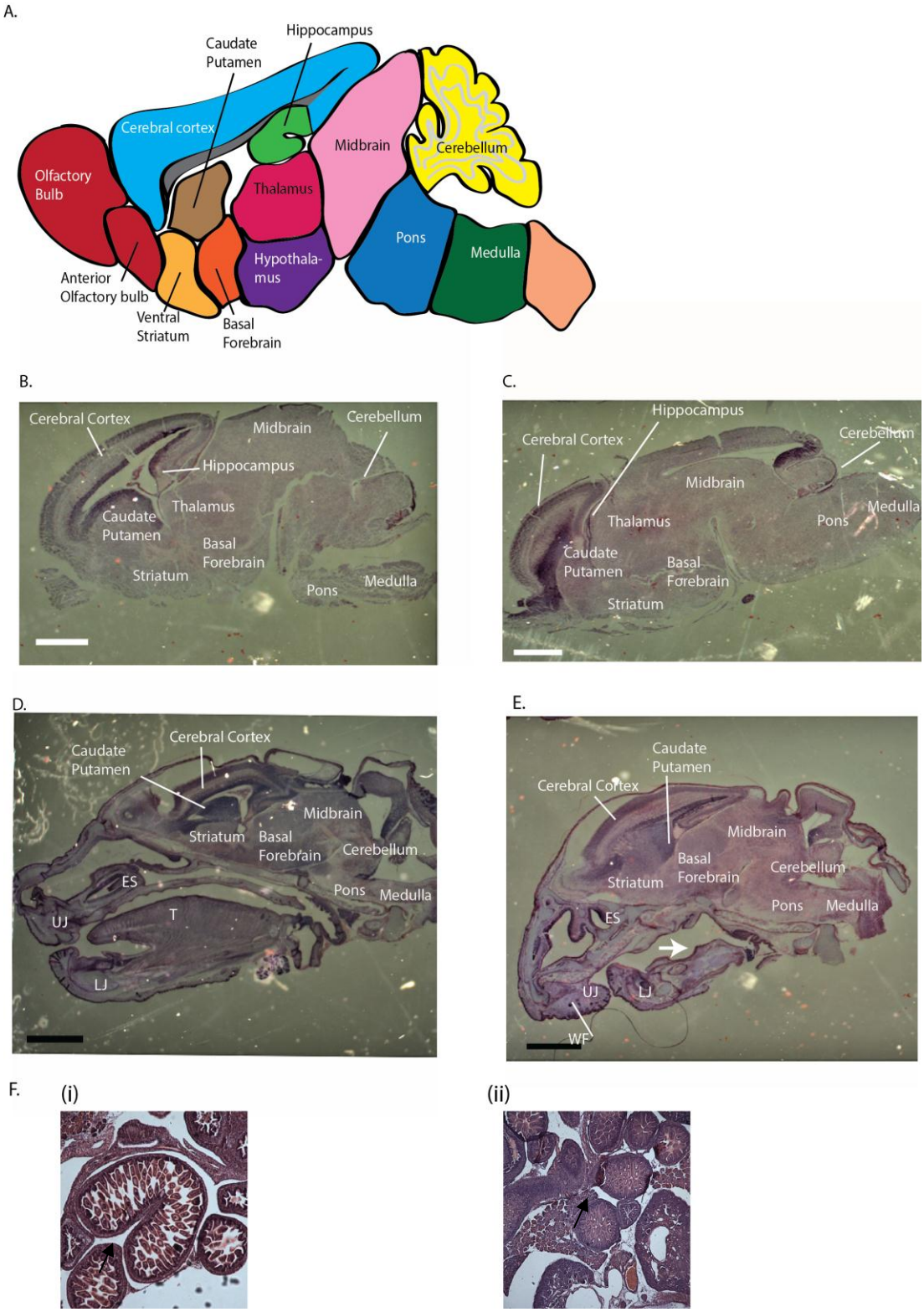


Figure 6.9: Histological analysis of the g/2-67 mouse line

A. Mouse brain anatomy in the wild type mouse, schematic shown above with H&E staining of sagittal sections through the brain of E17.5 wild-type embryo shown below. Brain structures labelled.

B. H&E staining of sagittal sections through the brain of E17.5 *g/2-67*$^{+/-}$*; Shh*$^{-/+}$ embryo exhibiting pointed face phenotype. Brain structures labelled, with no structures found to be missing.
C. H&E staining of sagittal sections through the skull of E17.5 wild-type embryo exhibiting normal development. Structures of the brain labelled and whisker buds (WB), cerebral cortex, striatum, cerebellum, eye socket (ES) tongue (T) and lower jaw (LJ) indicated.
D. H&E staining of sagittal sections across the skull of the E17.5 *g/2-67*$^{+/-}$*; Shh*$^{-/+}$ embryo exhibiting pointed face phenotype. Structures of the brain labelled and whisker buds (WB), cerebral cortex, eye socket (ES), lower jaw (LJ) indicated. The lower jaw is shown to be stunted and curve in under the upper jaw, a white arrow is also used to indicate the missing tongue.
E. (i) H&E staining of sections across the abdomen of an E17.5 wild-type embryo exhibiting normal development.
(ii)H&E staining of sections through the abdomen of E17.5 *g/2-67*$^{+/-}$*; Shh*$^{-/+}$ embryo exhibiting pointed face phenotype. Intestines shown to have dark staining between the intestinal villi (black arrow) as compared to the wild-type animal.
Scale bars used to show 100 microns.

Summary

The *g/2-67* deletion encompasses the *MRCS1* enhancer responsible for expression of *Shh* within the oral epithelium (Sagai et al., 2009) in addition it also removes the *MFCS4* enhancer which drives some expression within the oral epithelium and the palatal shelves of the soft palate (Sagai et al. 2009).

Within the developing face, mandibular processes give rise to the lower jaw skeleton and the pharyngeal epithelium influences the patterning and growth of the mandibular skeleton. The first pharyngeal arch (PA1) in mammals gives rise to jaws, skull wall, teeth, middle ear and part of the tongue, which is formed at E8.25. The arch rapidly increases in size as it is populated by mesenchyme derived from cranial neural crest cells (NCCs) (a population of highly migratory multipotent precursors that arise at the junction between the prospective neural tube and epidermis during early stages of vertebrate development (Gammill and Bronner-Fraser 2003) and develops into the mandibular and maxillary arches at E9.5. *Shh* is required for outgrowth of PA1 by regulatory epithelial-mesenchymal interactions that promote survival of mesenchymal cells. Reduction of *Shh* within the pharyngeal structures results in initial formation of a bilateral PA1 which then becomes hypoplastic resulting in a single fused structure in the midline by E9.5. Transgenic mice were utilised in order to produce NCCs lacking Shh, these were then shown to result in a

phenotype including a hypoplastic PA1 with midline or facial defects (Yamagishi et al. 2006). The mandibular arch is similar to limb buds in that both rely on interactions between mesenchyme and epithelium and both employ many of the same key regulatory genes (Schneider 1999). Regulation of limb size is believed to function via a signalling feedback loop involving *Shh* and *Fgf8*, in which cellular interactions are sensitive to distance within the growth field (Allard and Tabin 2009). It has been hypothesised that mandibular growth is regulated by a similar signalling feedback loop. *Shh* expression in the pharyngeal endoderm is essential for mandibular development, where it directs neural crest survival and outgrowth (Ahlgren and Bronner-Fraser 1999; Brito et al. 2006, 2008; Balczerski et al. 2012). Exogenous Shh, when placed in the mandibular arch environment, activates *Bmp4* and *Fgf8* expression in the oral ectoderm, resulting in the development of supernumerary mandibles (Brito, Teillet, and Le Douarin 2008). Development of the tongue requires complex epithelia-mesenchymal interactions in order to generate a structure capable of multiple biological functions. *Shh* plays a key role in the epithelial-mesenchymal interactions which are necessary for papilliogenesis. During early stages of embryogenesis, *Shh, Ptch1* and *Gli1* are expressed diffusely is the tongue (Hall et al. 1999). *Shh, Bmp2* and *Bmp4* expression is subsequently localised to regions of the anterior of the tongue where fungiform papilla form (Jung et al 1999). Inhibiting *Shh* with cyclopamine has been shown to result in enlarged papillae in ectopic regions (Jung et al. 1999).

Shh expression from the pharyngeal structures is therefore suggested to be important for lower jaw development. This implies that the pharyngeal enhancer would be important for development of the lower jaw; however, a lower jaw defect does not appear to result until deletion of both *MRCS1* and *MFCS4* has been undertaken as no facial defects are found within the g/2-120 line. Additionally *MFCS4Δ/MFCS4Δ* mice have previously been described within the literature and shown to exhibit defects of the epiglottis but not the face (Sagai et al. 2009). These data therefore suggests that the combination of the function of both these enhancers; *MFCS4* and *MRCS1*, are required to pattern the lower jaw. As both enhancers have been shown to be expressed within the both the palatal shelves and the tongue, it

suggests these structures are those which are affected by the g/2-67 deletion resulting in the observed phenotype.

6.3.6 LacZ reporter gene analysis of g/120

The g/120 deletion encompasses a 678 kb deletion including all three *Shh* epithelial enhancers as well as those which drive *Shh* expression within the forebrain. This deletion also brings the ZRS and *Shh* within 170 kb of each other (Figure 6.1). This deletion was confirmed as previously described by tail tip PCR.

Again, embryos were generated and harvested at stages E9.5-E14.5 to examine reporter gene expression (Figure 6.10). Forebrain expression within these animals was missing within embryos from E9.5, however, expression was still found to be present within the midbrain, hindbrain and the floor plate. At E11.5 the embryonic intestines were completely devoid of any *LacZ* expression, in contrast to g/2-120 and g/2-67 mice in which some degree of residual expression remained. This phenomenon was also seen within the guts of E14.5 day embryos thus suggesting an additional gut enhancer had been disturbed by this deletion, presumably *SLGE* which is located at close to the insertion site (Tsukiji, Amano, and Shiroishi 2013). The same was true of pharyngeal expression which was found to be undetectable. To date, no second enhancer has been reported for pharyngeal activity, these data however suggests it is possible that one is present (Figure 6.10). Embryos were also still found to have expression within the whisker buds and hair follicles as well as the anterior tongue within the older embryos. In addition the *g/120* mouse line (as well as the two smaller deletions) exhibited *LacZ* expression within the umbilicus. This again goes to suggest that yet unidentified enhancers still exist within the region and also that the expression found within the umbilicus is separable from that of the gut.

Despite *Shh* and the *Zrs* now being in such close proximity, limb expression still appeared to be normal within embryos at E9.5-E12.5. The g/120 embryos did appear to have a slightly higher level of limb bud expression at E13.5 compared to the SBLac control; however, this expression was not detectable by E14.5. The E13.5 embryos analysed were potentially a bit young, therefore this could be as a result of a

staging issue; repeating the staining at this stage would serve to confirm this. E11.5 embryos were dissected and were also found to exhibit a loss of notochord expression. Previously, it has been speculated that two notochord enhancers lie within 150 kb-600 kb from *Shh* (Jeong et al. 2006), although neither has so far been identified. Notochord expression was still found to be present within the g/2-67 line suggesting that the enhancers lie somewhere within the 430 kb between these two insertions (96-526 kb from *Shh*). SBE1 is responsible for the maintenance of *Shh* expression in the midline of the midbrain and caudal diencephalon from E10.5 onwards; is believed to act in concert with an additional enhancer which controls *Shh* transcription at earlier time points as well as driving expression within the Zli (See section 2.8). As no loss of midbrain or Zli expression occurs within the g/120 line it would suggest this additional enhancer is not located within the deleted region.

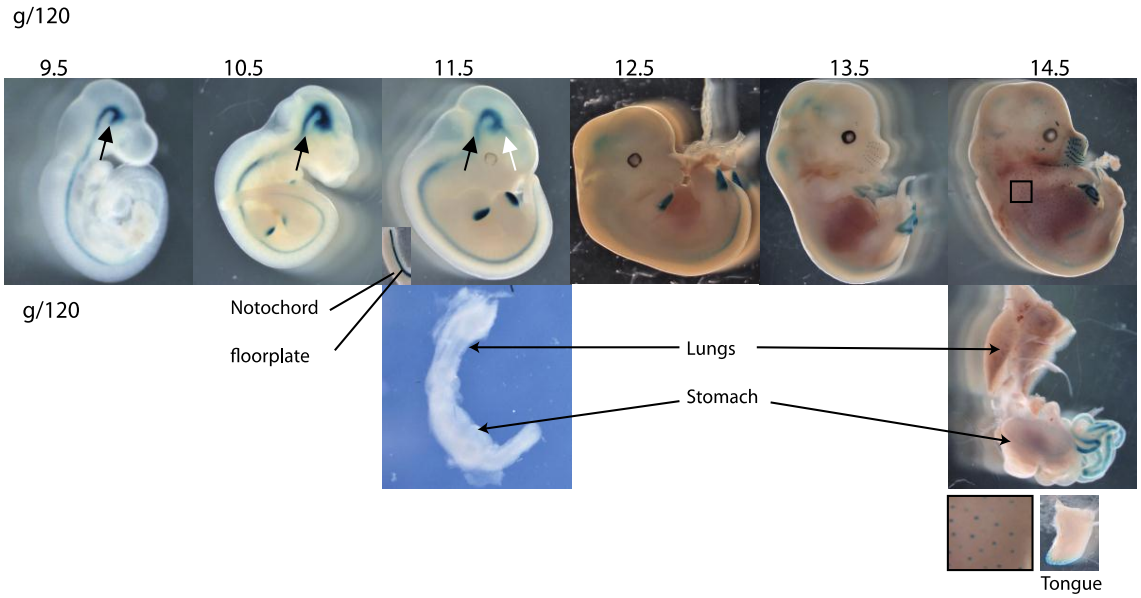


Figure 6.10: Reporter gene activity in the heterozygous g/120 mouse line

LacZ expression analysis of the g/120 deletion in heterozygous mice. Embryos stained from E9.5 to E14.5 including dissected guts, and neural tube at E11.5 and gut, skin and tongue at E14.5.Forebrain expression is shown (black arrow) to be missing from E9.5 onwards. No loss of Zli expression was found (white arrow). No expression is found within the gut at either stage and although floorplate expression is still present at E11.5 no notochord enhancer activity is detected.

6.3.7 Phenotype analysis within the g/120

Homozygous pups were identified for the g/120 line at E12.5 and examined for both *LacZ* reporter expression and phenotype (Figure 6.2). While heterozygous embryos at this stage (Figure 6.11A), showed normal development with *lacZ* reporter expression within the mid-brain, floorplate, limbs and whisker buds, homozygous g/120 embryos were found to have a highly variable phenotype. At the more severe end of the scale embryos were completely devoid of the facial structures, leaving only proboscis with one single eye located underneath it (Figure 6.11E). X-gal staining performed on these embryos demonstrated only staining within the forelimb, with all the structures of the brain, floor plate, notochord and gut devoid of stain. The phenotype exhibited in less affected embryos was found to be comprised of an upper jaw which appeared to have been split in half (Figure 6.11 B/C). Expression within these embryos was present within the midbrain and hindbrain as well as the floor plate and limb. However the forebrain *LacZ* expression was missing as was the notochord expression similarly to g/120 heterozygotes. Surprisingly staining within the gut (n=2) (Figure 6.11 D) and also the pharyngeal arches was present as was expression within the whisker buds although this was mis-expressed as a result of the malformation.

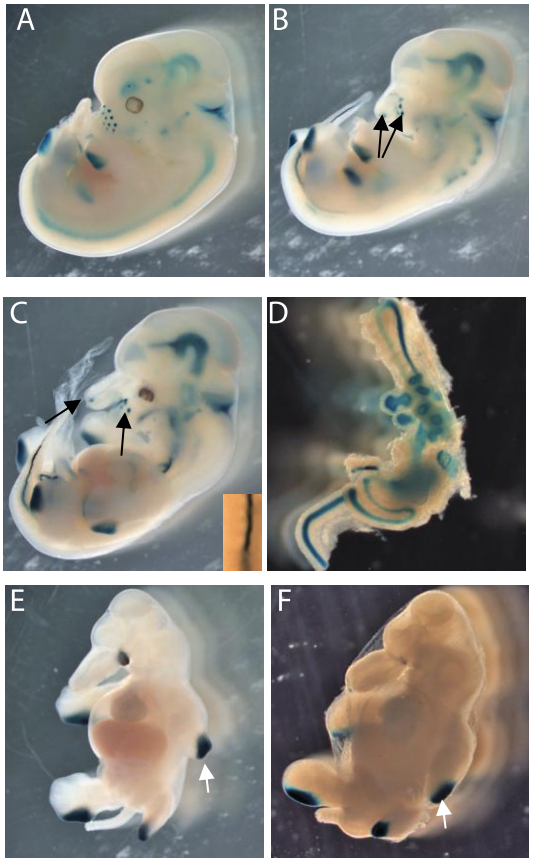


Figure 6.11: Phenotype analysis of the g/120 mouse line at E12.5

LacZ expression analysis within the g/120 homozygote 12.5 day embryos. (A) Normal development and staining pattern within *g/120* heterozygous mouse. (B/C) variations of moderate mutant morphology and staining pattern within *g/120* homozygotes. The upper jaws of moderately affected embryos shown to be split as demonstrated by the presence of whisker buds (black arrows) on both parts of the jaw. The embryo from (C) was dissected and loss of notochord enhancer although not floorplate shown in box within bottom right hand corner. The stained gut was also removed from this embryo (D). (E) More severe g/120 homozygous phenotype shown with LacZ reporter expression only found within the posterior of the limb bud (white arrow). (F) Comparable *g/120*$^{+/-}$*; Shh*$^{-/+}$ compound heterozygote shown at E12.5 again severe phenotype shown with staining only present within the limb buds (white arrow).

g/120$^{+/-}$*; Shh*$^{-/+}$ compound heterozygous embryos were also examined at E12.5 with animals shown to display a phenotype very similar to that of the severely affected homozygote (Figure 6.11F). These were stained with X-gal and again shown to have lost all but limb enhancer expression. E17.5 *g/120*$^{+/-}$*; Shh*$^{-/+}$ compound heterozygous embryos were also examined; similarly to E12.5 there was variation in the phenotype. At the milder end of the scale two embryos were found to be similar to *g/2-67*$^{+/-}$*; Shh*$^{-/+}$ mutants exhibiting a facial phenotype whereby the skull appeared

misshapen with a pointed snout and an absent tongue (Figure 6.12B[i]) as compared to a wild-type control (Figure 6.12A[i]). The lower jaw also appeared to be truncated within the *g/120*$^{+/-}$*; Shh*$^{-/+}$ animals. The brains of both the *g/120*$^{+/-}$*; Shh*$^{-/+}$ mutant as well as the control were removed and found to have no obvious morphological defect (Figure 6.12B [ii]) as compared to the wild-type (Figure 6.12A [ii]). The brains were also sectioned and again both g/120$^{+/-}$; Shh$^{-/+}$ mutants (Figure 6.12D) and wild-type (Figure 6.12C) appeared to be very similar. The guts of these animals were also examined by sectioning them and staining with H&E. Remarkably while the guts of both the g/2-120 and g/2-67 animals (Figure 6.6/6.9) were morphologically abnormal this did not appear to be the case with the mild *g/120*$^{+/-}$ *; Shh*$^{-/+}$ compound heterozygote animal (Figure 6.12B [iii]). Instead the gut of this animal was found to resemble the wild-type control (Figure 6.12A [iii]).

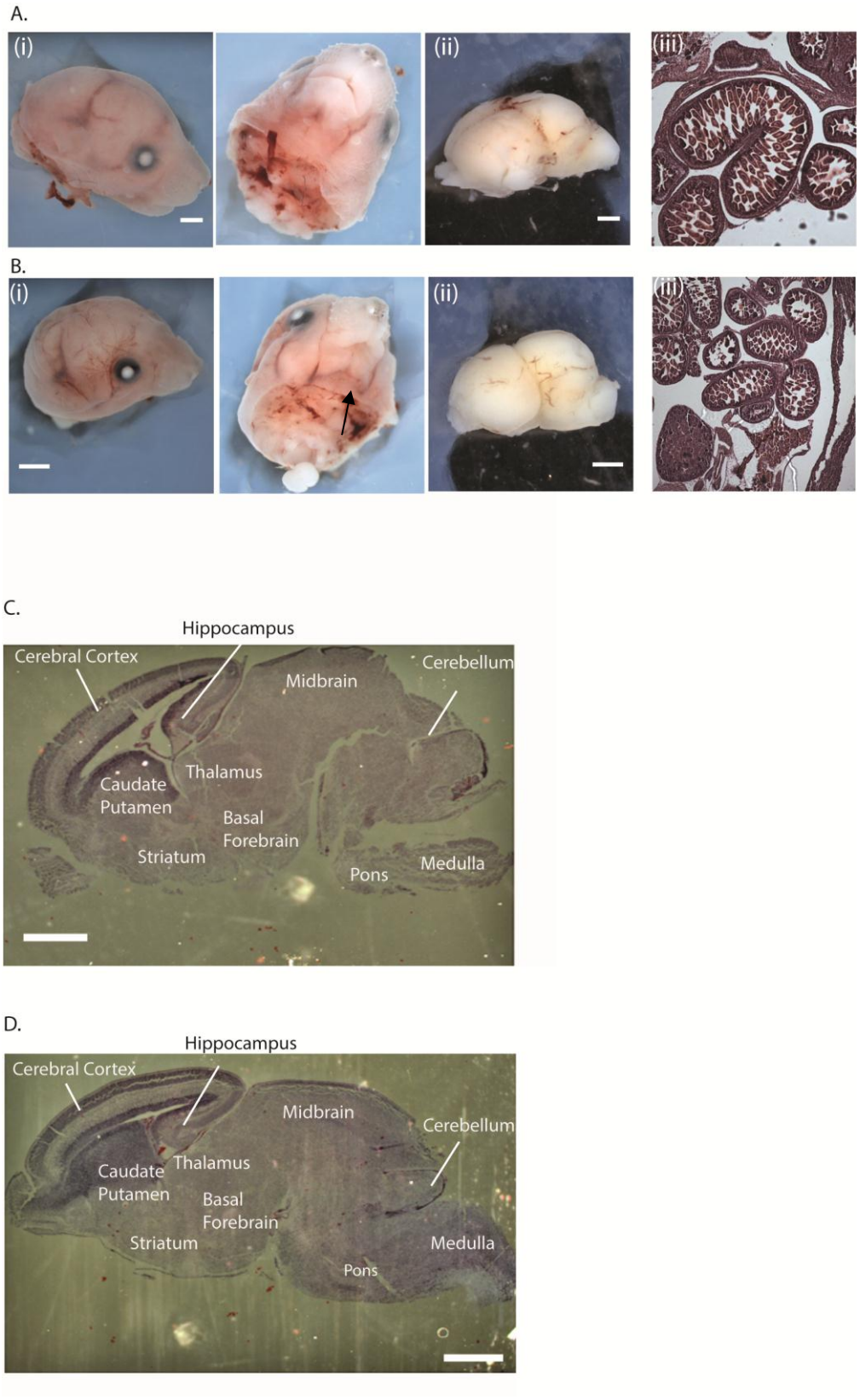


Figure 6.12:Phenotype analysis of mildly affected g/120 mice at E17.5

A. (i) Dissected head from E17.5 day wild type embryo exhibiting normal development. (ii) Dissected brain from E17.5 day wild type embryo (iii) H&E stained section taken from sagittal section through the abdomen of wild-type E17.5 day embryo.
B. (i) Dissected head from *g/120*$^{+/-}$*; Shh*$^{-/+}$ E17.5 day embryo exhibiting a mild phenotype. The head is shown to be pointed, with a truncated lower jaw which does not appear to have fused (black arrow) and missing tongue.
(ii) Dissected brain from *g/120*$^{+/-}$*; Shh*$^{-/+}$ E17.5 day embryo exhibiting mild facial phenotype. Brain was shown to look somewhat morphologically abnormal.
(iii) H&E stained section taken from sagittal section through the abdomen of *g/120*$^{+/-}$*; Shh*$^{-/+}$ E17.5 day embryo exhibiting mild phenotype. No differences found with respect to the control.
C. H&E stained section taken from sagittal dissection of a brain from of a wild-type E17.5 day embryo. Brain structures labelled and indicated with black line.
D. H&E stained section taken from sagittal dissection of brain of *g/120*$^{+/-}$*; Shh*$^{-/+}$ E17.5 day embryo exhibiting mild facial phenotype. Brain structures present (labelled and indicated with black line).
Scale bars used to show 100 microns.

A *g/120*$^{+/-}$*; Shh*$^{-/+}$ compound heterozygote was also identified at E17.5 exhibiting a moderate phenotype. This comprised a very obvious facial phenotype; this time the animal's snout was found to resemble a split upper jaw whilst the lower jaw was also severely truncated suggesting the pharyngeal arches had not fused. This embryo was much smaller than its littermates with respect to body and head size and also exhibited a loss of the eyes and its ears also appeared abnormally placed (Figure 6.13A). Sectioning of the head suggested that although an eye socket was present the eye itself was not; the tongue was also shown to be absent. Although neural tissue was present within the brain of the animal it was not patterned as expected. The major difference found within the brains of the *g/120*$^{+/-}$*; Shh*$^{-/+}$ animals was that the striatum was missing (See Figure 6.9A), the cerebral cortex as well as the hippocampus were present although they had collapsed without the support of the ventral brain tissues (personal communication Professor David Price) (Figure 6.13B). The guts within this animal were also sectioned, and an obvious gut defect was identified (Figure 6.13C [ii]). However, unlike the two previous smaller deletions where the guts were shown to have increased amount of tissue within the villi with respect to the control (Figure 6.13C [ii]) these animals were in fact found to have an less tissue within the intestine, which was a similar phenotype to that found within the *Shh* null mice (Figure 6.13E [ii]).

The final and most severe *g/120*$^{+/-}$*; Shh*$^{-/+}$ mutant found was one which was most similar to the *Shh* null (Figure 6.13E[i]) with the only difference being this embryo

had formed legs and digits (Figure 6.13E[i] F), appearing to be of normal size, although twisted potentially as a result of the reduced size of the embryo. Most of the head and facial features were missing and the forebrain was found to represent an elongated proboscis with a single eye located at the end. Several of the *g/120*$^{+/-}$*;-Shh*$^{-/+}$ embryos were also found to exhibit a curled tail and a kinked neural tube presumably as a result of loss of the notochord enhancer. The gut from the severe g/120$^{+/-}$; Shh$^{-/+}$ embryo was also sectioned and stained with H&E (Figure 6.13D [ii]), this had a phenotype very similar to both the moderately affected *g/120*$^{+/-}$*; Shh*$^{-/+}$ embryo as well as the *Shh* null (Figure 6.13E [ii]); whereby, an increased amount of space was found within the intestinal epithelium.

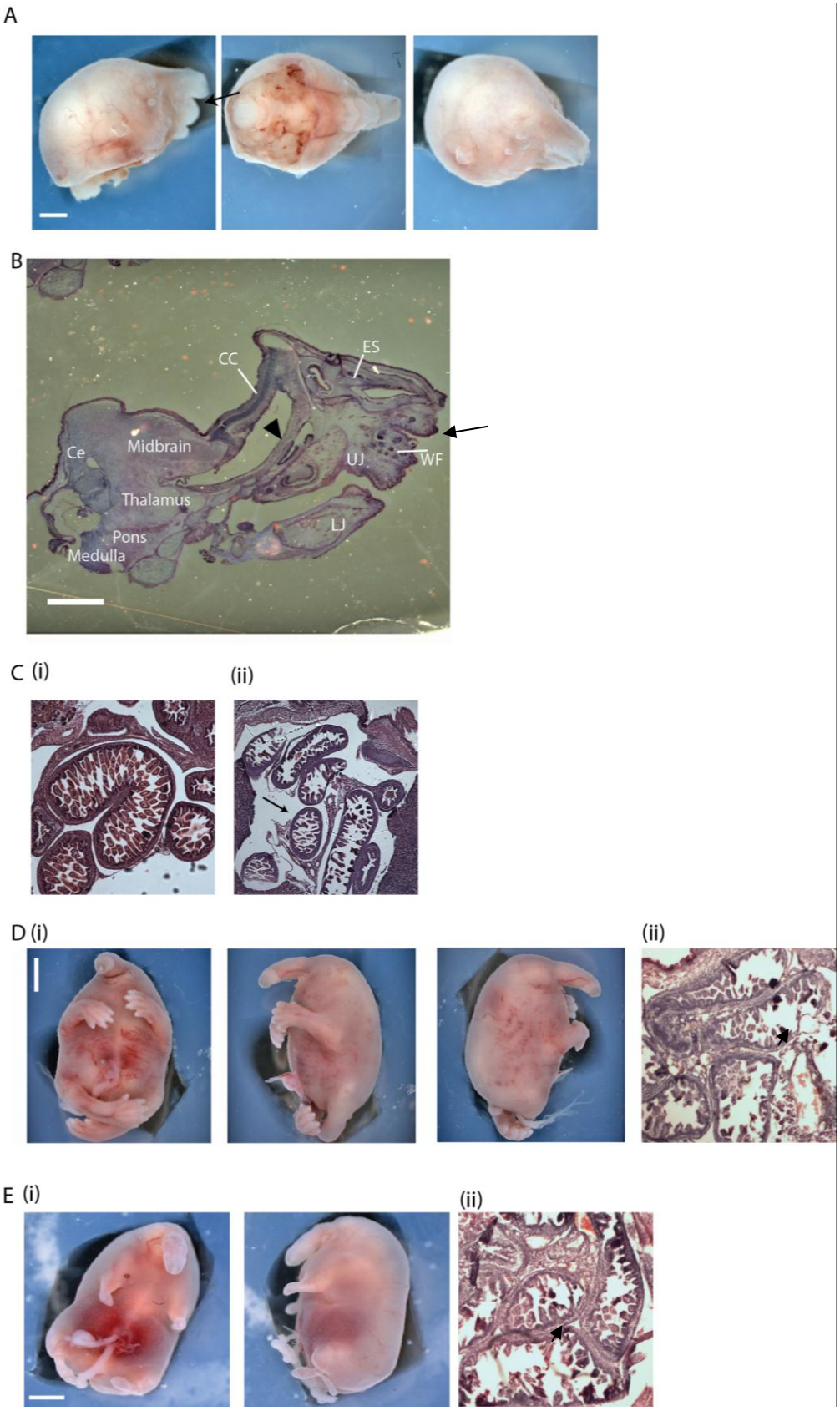


Figure 6.13: Phenotype analysis of moderate/severely affected g/120 mice at E17.5

A. Dissected head from *g/120*$^{+/-}$*; Shh*$^{-/+}$ E17.5 day's embryo exhibiting moderate phenotype. Split upper jaw indicated by black arrow.
B. H&E stained section taken from sagittal dissection of skull from *g/120*$^{+/-}$*; Shh*$^{-/+}$ E17.5 day embryo moderate facial phenotype as shown in (A). Cerebellum (Ce), cerebral cortex (CC), eye socket (ES), whisker buds (WB), upper jaw (UJ) and lower jaw (LJ) all designated. Split upper jaw/proboscis indicated by black arrow. Missing forebrain structures highlighted by black arrow head.
C. (i) H&E stained section taken from sagittal dissection of abdomen of a wild-type E17.5 day embryo exhibiting normal development. (ii) H&E stained section taken from sagittal dissection of abdomen of *g/120*$^{+/-}$*; Shh*$^{-/+}$ E17.5 day embryo exhibiting moderate phenotype. Black arrow indicates decreased tissue with regards morphology as compared to wild-type embryo.
D. (i) *g/120*$^{+/-}$*; Shh*$^{-/+}$ embryo at E17.5 exhibiting severe phenotype including craniofacial defects. (ii) H&E stained section of abdomen of *g/120*$^{+/-}$*; Shh*$^{-/+}$ E17.5 day embryo exhibiting severe phenotype presence of decreased tissue indicated by black arrow.
E. (i) *Shh* null embryo at E17.5 exhibiting craniofacial phenotype and truncated limbs. (ii) H&E stained section taken from abdomen of *Shh* $^{-/-}$ E17.5 day embryo presence of increased space indicated by black arrow.
F. *g/120*$^{+/-}$*; Shh*$^{-/+}$ embryo at E17.5 (left) versus *Shh* null embryo at E17.5 (right) showing size difference between embryos.
Scale bars used to show 100 microns.

Summary

The facial defects exhibited by the g/120$^{+/-}$; Shh$^{-/+}$ mouse mutant could be attributed to the loss of the oral epithelial enhancer. However; as both, the severe and moderate phenotypes identified were more severe than the *g/2-67*$^{+/-}$*; Shh*$^{-/+}$ phenotype it would suggest that forebrain expression of Shh is also important for the development of the face. Data from HPE3 patients who carry translocations with breakpoints 250 kb upstream of *SHH* have been shown to exhibit forebrain and mid-face defects including a mild HPE craniofacial defect, aplasia of the premaxilla, hypotelorism, midline cleft lip and hearing loss (Belloni et al. 1996). *SBE2* is found 250 kb from *SHH* which would therefore be disrupted by such a breakpoint.

The embryonic telencephalon, which is located at the most rostral end of the neural tube, is divided into the dorsal telencephalon, which gives rise to the neocortex, and the ventral telencephalon, which forms the striatum and is the origin of cells that populate the olfactory bulb, globus pallidus, and some cells that also populate the cortex (Evans et al 2012). Shh is first secreted from the notochord, which underlies the posterior structures of the brain (Echelard et al. 1993; Roelink et al. 1995; Ye et al. 1998). Shh operates through a concentration gradient that spans the DV axis of the brain at different time points to confer different neuronal identities on the developing precursors, and expression is first seen in the ventral telencephalon

from E11.5 (Kohtz et al. 1998; Jessell 2000). *Shh* expression directs neural progenitors to a ventral fate and importantly is both necessary and sufficient to induce specific ventral forebrain markers (Chiang et al. 1996). Shh is expressed in the ventral telencephalon where it interacts with the Gli proteins and is thought to maintain Fgf8 expression as well as induce expression of forebrain markers (Ohkubo et al. 2002).

The inhibition of the Gli3 repressor complex in the ventral region facilities telencephalon development; therefore, the primary function of Shh is to prevent the production of excessive Gli repressor. In *Shh* null mutants, the ventral markers *Dlx2*, and *Gsx2* are reduced, whereas in *Gli3* null mouse mutants the expression pattern of these genes is extended dorsally (Rallu et al. 2002). *Shh* null mutants show more severe telencephalon abnormalities than the *Gli3* null mice (Grove et al. 1998; Theil et al. 1999). In *Shh* mutants, there is a loss of ventral telencephalic cells leading to an altered morphology of the ventral telencephalon together with the ectopic expression of dorsal forebrain markers (Chiang et al. 1996; Rallu et al. 2002). However, if *Shh* is knocked-out at E8.5, there are severe defects of all ventral telencephalic regions (Fuccillo et al. 2004). However, in knockouts at E10-12 using a Nestin-Cre, there are limited defects in ventral telencephalic patterning and cortical interneurons are affected rather than gross patterning deficits (Xu et al. 2005). Therefore between E9 and E12.5, Shh acts mainly by inhibiting the formation of the Gli3 repressor and contributes to the establishment of DV patterning (Chiang et al. 1996; Rallu et al. 2002; Fuccillo et al. 2004). Secondly, SHH signalling also supports the expansion of progenitors of the ventral telencephalon by inducing and maintaining the expression of *Nkx2.1* until at least E14 and later into neurogenesis (Xu, Wonders, and Anderson 2005).

Shh is produced by multiple epithelia in the head including forebrain neuro-ectoderm, frontonasal and maxillary ectoderm and pharyngeal endoderm (Marcucio et al. 2005). Studies conducted in mammals and birds suggest that the severity of HPE defects correlate with the stage in which interruption of signalling occurs. Null *Shh* mutations within the mouse disrupt neural plate patterning and cause cyclopia. Disruption of *Shh* expression by the teratogen cyclopamine (See section 2.6) results

in a varied phenotype, with the time-point at which *Shh* signalling response is lost being pivotal for phenotype. If cyclopamine exposure takes place before *Shh* is established in the telencephalon, embryos had a single fused telencephalic vesicle and also exhibited extreme hypotelorism. In these cases the facial defects arose as a consequence of the underlying structural alterations in forebrain architecture. By contrast, if cyclopamine was delivered after *Shh* expression was established in the forebrain but before its induction in the face, embryos had two well-delineated telencephalic vesicles that were morphologically and histologically normal. However despite the normal appearing brains the embryo was still shown to exhibit severe hypotelorism and were found to down-regulate target genes in the adjacent mesenchyme, which leads to a narrow and truncated face as a result of patterning defects (Marcucio et al. 2005; Chong et al. 2012).

NCCs are also likely to play a role; their diverse fates include several cell types within the peripheral nervous system, melanocytes, endocrine cells, and most of the bone, cartilage, and connective tissue of the face and skull (Santagati and Rijli 2003). NCCs that contribute to the face originate from the caudal forebrain, midbrain, and rostral hindbrain. These cranial neural crest cells (CNCCs) follow well defined paths to the ventrolateral side of the head, where they populate the mesenchyme of the facial primordia such as the frontonasal prominence (FNP) and pharyngeal arch; (Osumi-Yamashita et al. 1994; Kontges and Lumsden 1996). Once positioned, the NCCs proliferate and differentiate into distinct craniofacial elements (Helms and Schneider 2003; Santagati and Rijli 2003). Mice in which *Shh* has been inactivated within the NCC appear indistinguishable from the wild type embryos until E10.5, when they have slightly smaller FNPs and mandibular lower jaw. The growth deficiency of the *Shh* NCC mutant facial primordia is very obvious by E12.5, and as a result, mice are born with a dramatically truncated face and also lack a tongue (Jeong et al. 2004). It is suggested that the role of SHH within NCC is to support cell survival during early stages and promotes proliferation at later stages to control the size of facial primordia (Jeong et al. 2004).

Expression of *Shh* within the notochord has been shown to be important for patterning of the floorplate, hind-brain and mid-brain (See section 2.3). Experiments have also shown that if the notochord is separated from midbrain floor plate during development, the brain vesicles exhibit abnormal folding and collapse (Britto, Tannahill, and Keynes 2002). The g/120 mouse line has been shown by reporter gene analysis to result in a loss of expression within the notochord at all stages examined.

Expression of *Shh* and its many roles in the development of both the face and brain is therefore complex. The variable phenotype seen within the g/120 mouse could therefore be as a result of various factors. Temporal changes may play a role whereby embryos in which *Shh* has been lost from the forebrain earlier may result in the more severe forebrain phenotype, whereas, those where the expression has been lost later result in milder defects. Reporter gene analysis suggests that LacZ expression is lost from E9.5 in the g/120 heterozygote (Figure 6.10). Whereas staining of the homozygote embryos, even with the milder phenotype suggests forebrain expression appears to be missing from E12.5 days (Figure 6.11) (younger embryos were not examined) suggesting no temporal changes were found with regards to loss of enhancer activity. Alternatively the variation in forebrain phenotype could therefore be a result of genetic rescue within the milder phenotype. How this could be occurring remains unclear; Shh is known to be a diffusible morphogen capable of working at large distances. Shh has also been shown to emanate from various structures including the Henson's node and pattern the forebrain. Ihh has been suggested to be capable of rescuing deleterious phenotypes within the CNS (Wijgerde et al. 2002). Additionally, the HPE phenotype results in a broad spectrum of clinical phenotypes observed in families with identical mutations. This variation in has been attributed to there being a threshold of SHH protein; below this threshold causes a phenotype, above does not (Jeong et al. 2008). It could therefore be proposed that Hh signalling from other tissues, especially if the threshold levels are low could be responsible for rescuing the forebrain defect.

The variation in phenotype could also be attributed to result from the compound effect of the loss of notochord and forebrain Shh expression. The milder phenotypes could, therefore, be as a result of only forebrain disruption; whereas, the more severe phenotypes as a result of disruption of SHH expression in both tissues leads to the collapse of the structures of the brain and floorplate. Reporter gene expression analysis of both compound heterozygous and homozygous embryos at E12.5 could support this theory (Figure 6.11). Within the more severely affected embryos, enhancer expression within tissues such as the hindbrain, mid-brain and floorplate were also missing. Due to the severity of the phenotype within these embryos this was hypothesised to be as a result of the failure of these tissues to form; i.e., if the tissues did not develop it would not be possible to detect LacZ expression within them. This theory also suggests that rescue of tissues by very low level SHH protein production is likely to be occurring within tissues of milder affected embryos.

Although g/120 heterozygous embryos were shown to lose reporter gene expression within the gut and pharyngeal arches, the homozygotes with milder phenotypes were actually found to have LacZ expression within these tissues (Figure 6.10). The phenotype within the mildly affected E17.5 $g/120^{+/-}; Shh^{-/+}$ embryos were also shown to have normal gut morphology (Figure 6.12) as opposed to both the g/2-120 (Figure 6.13) and g/2-67 lines (Figure 6.9). This suggests some form of mechanism exists whereby enhancer function can be rescued. Critical developmental control genes sometimes contain ‘shadow’’ enhancers that can be located at large distances and in locations, including the introns of neighbouring genes. They nonetheless produce patterns of gene expression that are the same as or similar to those produced by primary enhancers. Gene expression levels naturally fluctuate between individuals in a population and between cells in a tissue. Genetic variation and environmental instability can exacerbate these differences. If the expression of a developmental regulator dips below a minimum threshold level, cell fate specification or differentiation will be affected, which can result in a morphological defect. Developmental programs can be buffered against minor variations by maintaining super-threshold expression levels of important regulators. Shadow enhancers help foster robustness in gene expression in response to environmental or genetic perturbations (Frankel et al. 2010; Hobert 2010; Perry et al. 2010). The *SLGE*

enhancer which had recently been identified is believed to represent one such shadow enhancer which works in combination with *MACS1* to drive expression within the gut. The g/2-120 versus g/120 deletion lines suggests that loss of *MACS1* but not *SLGE* prevents a complete loss of reporter gene activity within the gut as well as more severe gut defects (Figure 6.7/6.10). It is therefore entirely possible that additional shadow enhancers also exist which are capable of rescuing *Shh* expression in a number of tissues.

6.4 How does the presence of multiple promoters affect enhancer activity?

Within the developing embryo enhancers are rarely permitted promiscuously to activate transcription from non-target promoters. Chromatin structural organization and insulators are thought to play a role in regulating enhancer-promoter interactions (Celniker and Drewell 2007; Wallace and Felsenfeld 2007; Bushey et al. 2008). Previous work (see section 5.11) also suggests that the promoters of certain genes such as *Rnf32* and *Rbm33* present within the cis-regulatory regions of *Shh* and *En-2* are refractory to enhancer activity. Positive and negative control regions have been shown to be shared between promoters within adjacent *Hox* genes (Sharpe et al. 1998). Enhancer elements have also been shown at the HoxD locus to activate neighbouring genes, a so called 'bystander effect' (Spitz, Gonzalez, and Duboule 2003). This is believed to be either as a consequence of chromatin activation of a region or alternatively due to 'promoter competition' providing a way to filter out some enhancer activity from the target promoter. Various examples of loss or gain of expression have been reported following transgene insertion at the proximity of regulatory elements (Olson et al. 1996; Sharpe et al. 1999; Qin et al. 2004; Aubert et al. 2010) potentially caused by complex titration-like mechanisms of either positive or negative regulations.

The LHED system which was utilised to distribute the *LacZ* reporter gene throughout the *Shh* region provided a means to examine whether the enhancers

surrounding *Shh* are firstly capable of activating more than one promoter within a tissue and also whether any form of promoter competition exists.

The SBLac120 mouse line contains the SBLac insertion approximately 100 kb from *Shh* prior to Cre recombinase-generated deletion. SBLac120 mice were crossed to generate mice homozygous for the insertion i.e. had a *LacZ* reporter complex including a HSP68 promoter located 100 kb from *Shh* (Figure 6.14A). Generating these homozygotes would enable investigation of whether enhancers within the region can activate both the *LacZ* reporter gene and the *Shh* gene within the same tissue, determined by whether development of these tissues remains normal despite activation of LacZ.

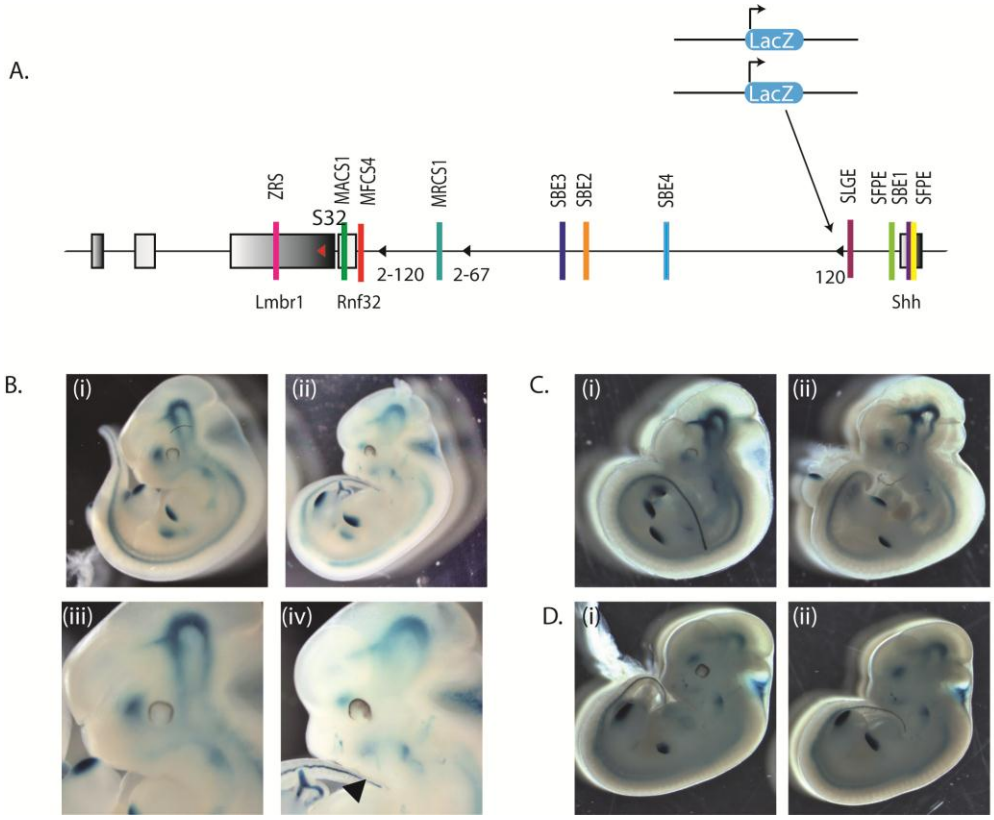


Figure 6.14: The effect of multiple promoters within a genomic locus

A. Schematic diagram illustrating the presence of the LacZ reporter genes, one on each chromosome (blue rounded rectangle) including promoter inserted 100 kb from *Shh*. Presence of various enhancers designated by coloured rectangles and genes by grey rectangles. Insertion sites of the three SBLac lines investigated shown by black arrow heads.
B. (i) LacZ expression analysis of E11.5 SBLac120 embryos within heterozygous mice. Expression found within the CNS, gut epithelium and posterior limb bud (ii) LacZ expression

analysis of E11.5 SBLac120 embryos within homozygous mice, missing expression from the forebrain with abnormal expression found within the facial structures and a facial defect also observed (iii) Developing pharyngeal arches within the SBLac120 heterozygous embryo shown in (i). Development appears to be normal. (iv) Developing pharyngeal arches within the SBLac120 homozygous embryo shown in (ii) fusion of the pharyngeal arches has not occurred (black arrow head).

C. (i) LacZ expression analysis of E11.5 SBLac2-120 heterozygous embryos exhibiting normal expression pattern found in the CNS, gut and limb. (ii) LacZ expression analysis of E11.5 SBLac2-120 homozygous embryos also found to exhibit a normal expression pattern.
D. (i) LacZ expression analysis of E12 SBLac2-67 heterozygous embryos, exhibiting normal expression pattern in the CNS, gut and limb. (ii) LacZ expression analysis of E12 SBLac2-67 homozygous embryos exhibiting normal expression pattern.

No SBLac120 homozygotes were found after birth suggesting that the phenotype generated could be lethal (Figure 6.2). Embryos were therefore harvested at both E11.5 days and E17.5 days, with SBLac120 homozygotes identified at these time points. At E11.5 days embryos were assessed for phenotype and also stained with X-gal in order to analyse *LacZ* reporter expression. At 11.5 days LacZ reporter expression was found throughout the CNS, gut and limb of heterozygous SBLac120 embryos and no abnormal phenotypes were observed (Figure 6.14B [i/iii]). In contrast, homozygous SBLac120 embryos were found to have malformed facial development with a somewhat shortened face and also non-fusion of the pharyngeal arches, the rest of the embryo including the limb however appeared normal (Figure 6.14B [ii/iv]). With respect to *LacZ* expression, despite having two copies of the *LacZ* reporter gene the overall expression pattern (with the exception of the limb bud) appeared to be less intense than the heterozygote carrying one copy. A normal expression pattern equivalent to the heterozygote was found within the neural tube as well as the limbs, gut and midbrain. The forebrain expression was missing, whilst within the pharyngeal structures mis-expression appeared to have occurred. Embryos were also generated in order to examine homozygous embryos from both the SBLac2-120 (Figure 6.14C) and SBLac2-67 (Figure 6.14D) lines so as to examine the effect of multiple promoters at other regions within the locus. These embryos were harvested at approximately E11.5 and stained using X-gal to look at *LacZ* reporter gene expression. Surprisingly the expression pattern did not appear to differ between the homozygotes (Figure 6.14C/D [ii]) and heterozygotes (Figure 6.14C/D [i]). No obvious phenotype was found to have occurred either.

At E17.5 days SBLac120 homozygous embryos were again shown to have variable cranio-facial phenotypes as compared to the SBLac120 heterozygous embryos (Figure 6.15A). Homozygous embryos had a reduced bottom jaw with or without the presence of a tongue (Figure 6.15 B/C). In addition, homozygotes appeared to have shortened snouts, as well as a somewhat pointed faces as compared to heterozygotes, both phenotypes previously attributed to loss of late Shh forebrain expression (Marcucio et al. 2005; Chong et al. 2012). The heads from the homozygous embryos were sectioned and staining with H&E was undertaken. From this it was shown that both the upper and lower jaw was truncated, with the tongue present within one embryo but not the other (Figure 6.15 B/C [iii]). The forebrain was also disrupted as compared to a wild-type animal (Figure 6.14 A [iii]) with the cerebral cortex being present, however, with no structures of the forebrain such as the striatum present underneath (Figure 6.15 C/D [iii]) (See Figure 6.9A). The rest of the embryo appeared phenotypically normally with no outwardly adverse effects to the gut or limb exhibited. The guts were sectioned and stained with H&E (Figure 6.15 D (ii)) and found to be comparable to those of a wild-type control (Figure 6.15 D (i)). Overall the craniofacial defects exhibited within the SBLac120 homozygotes would suggest that these embryos were suffering from defects of the forebrain enhancers (*SBE2-4*) as well as the pharyngeal (*MFSC4*).

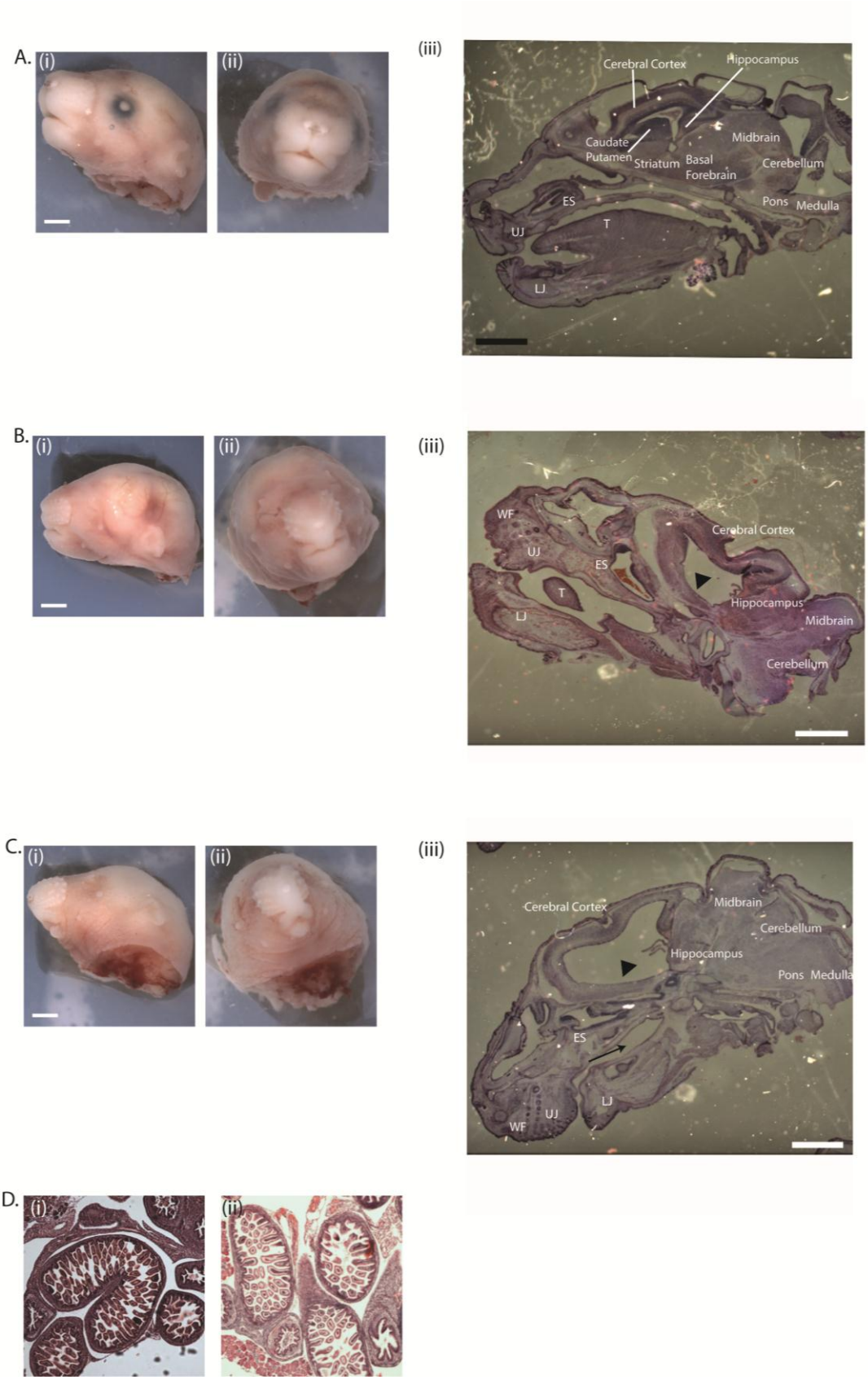


Figure 6.15: Phenotype analysis of SBLac120 homozygous mice at E17.5

A. Front (i) and side view (ii) of E17.5 head of a wild-type mouse showing normal cranio-facial development. (iii) H&E stained section taken from sagittal section of skull from a wild-type E17.5 day embryo exhibiting normal facial development. Eye socket (ES), tongue (T), upper jaw (UJ) and lower jaw (LJ) all designated and brain structures indicated.

B. (i) Front and (ii) side view of E17.5 head of SBLac120 homozygote mouse showing one presentation of abnormal cranio-facial development including shortened/pointed face and shortened lower and upper jaw (iii) H&E stained section taken from sagittal section of skull from SBLac120 homozygous E17.5 day embryo shown in (i/ii) exhibiting facial phenotype. Eye socket (ES), Tongue (T), whisker b (WF), upper jaw (UJ) and lower jaw (LJ) all designated and brain structures indicated. Black arrow head used to indicate the loss of forebrain structures such as the striatum.

C. (i) Front and (ii) side view of E17.5 head of SBLac120 homozygote mouse showing second example of abnormal cranio-facial development including shortened/pointed face and absent or missing lower jaw (iii) H&E stained section taken from sagittal section of skull from SBLac120 homozygous E17.5 day embryo shown in (i/ii) exhibiting facial phenotype. Eye socket (ES), Tongue (T), whisker follicle (WF), upper jaw (UJ) and lower jaw (LJ) all designated and brain structures indicated. Black arrow head used to indicate the loss of forebrain structures such as the striatum. Black arrow indicates the loss of the tongue.

D. (i) H&E stained section taken from a section of abdomen of a wild-type E17.5 day embryo exhibiting normal development. (ii) H&E stained section taken from a section of abdomen of a SBLac120 homozygous E17.5 day embryo exhibiting facial defect but no gut abnormalities.

Scale bars reflection 100 microns shown.

6.5 Conclusions

6.5.1 Expression analysis

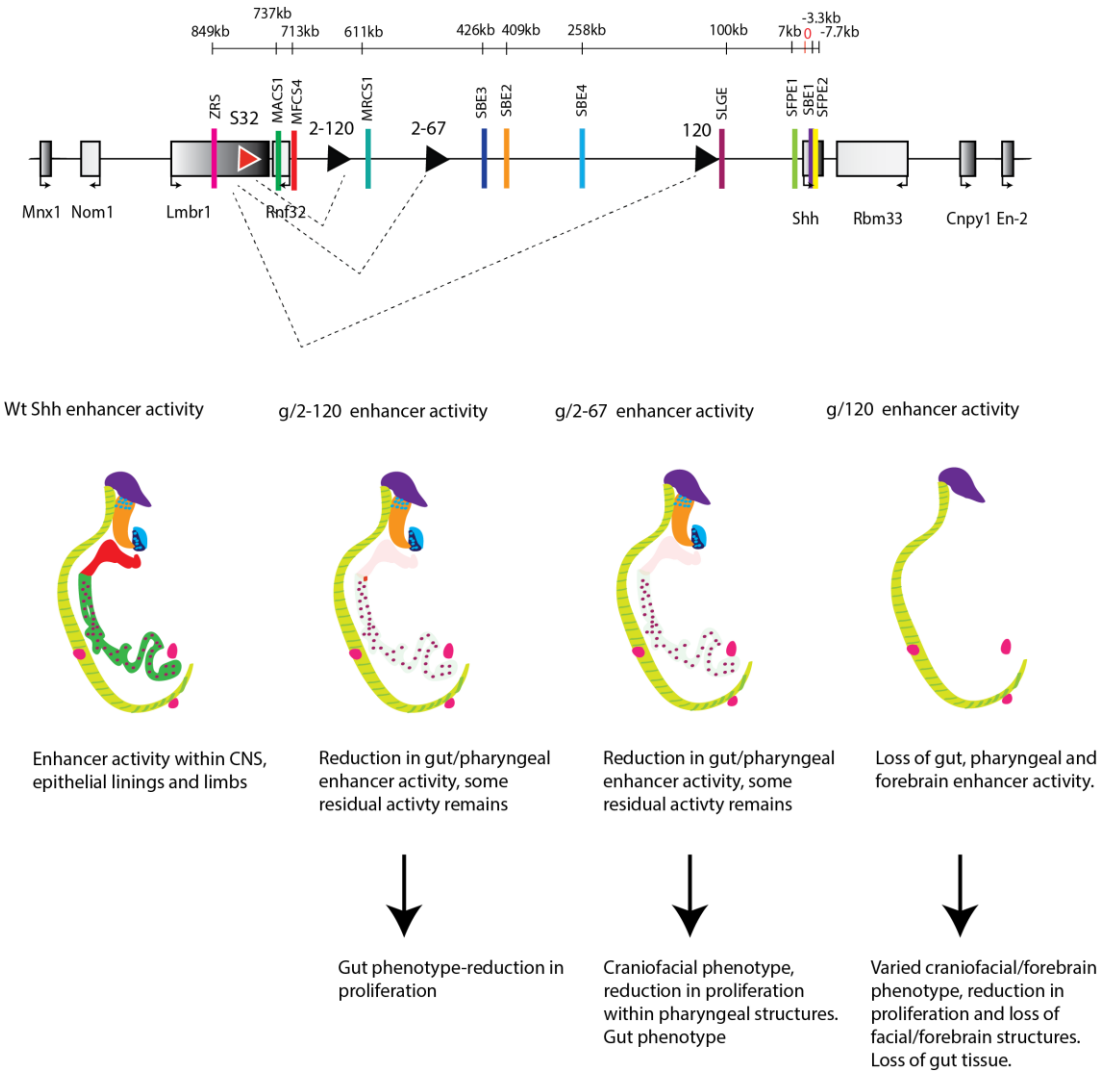


Figure 6.16: How do deletions within the *Shh* region affect reporter gene expression and phenotype?

Top: Shh locus shown with each enhancer represented by a coloured bar, location of the original LHED insertion site, S32, shown as a red triangle. The location of the three re-insertions used to generate the lines g/2-120, g/2-67 and 6/120 represented by black triangles and the region deleted within these lines shown by dotted line. **Bottom**: Activity of the Shh enhancers within the wild-type embryo shown, with the tissues of activity colour coordinated with the enhancers shown above. Expression found within the CNS, epithelial linings and the limbs. Enhancer activity following each of the deletions showed which results in loss or reduction of enhancer activity within various tissues. Phenotypes generated in each of these lines also stated.

All three deletion lines were shown to result in changes to expression across the embryonic time course; however, none appeared to affect the temporal expression pattern but instead the spatial pattern only, with one exception being the limb buds within the g/120 line which needs to be further examined. Each expression pattern was shown to reflect that of a deletion of the corresponding enhancer with the g/2-120 line resulting in loss of gut expression, and g/120 resulting in loss of gut and forebrain expression. Unexpectedly however, loss of the oral epithelial enhancer within SBLac2-67 (and SBLac120) did not appear to effect whisker bud expression as hypothesised. The expression patterns also demonstrated the modular nature of the *Shh* enhancers. A number of *Shh* enhancers have already been identified which also include secondary enhancers responsible for driving expression in the various tissues. The three deletions generated within this study show that deletion of one enhancer does not affect the activity of the remaining enhancers in the region i.e. it is possible to remove the gut and forebrain enhancers without disrupting activity of the *Zrs* or mid-brain enhancer *SBE1*(Figure 6.16). This suggests that the *Shh* region is not behaving in the nature of an archipelago or holoenhancer such as that reported for other developmental regulatory landscapes within the literature (Montavon et al. 2011; Marinic et al. 2013) but instead each enhancer is independent of the others, with even secondary enhancers playing independent roles. The order of the *Shh* enhancers was established in early vertebrate evolution and is highly conserved (Hadzhiev et al. 2007). It has been suggested that this conservation means that this conserved landscape is important for enhancer function; however this seems unlikely to be the case. Deletions of up to 700 kb of DNA from this region do not appear to affect how the remaining enhancers behave. The *Zrs*, in particular, is located nearly 1 Mb away from *Shh* and it was believed that this distance could be important to function; however, reducing the distance does not appear to result in major changes to reporter gene expression (Figure 6.16).

The gut enhancer *MACS1* and pharyngeal enhancer *MFCS4* were deleted within all three lines however a complete loss of reporter expression within the gut was not seen until the largest deletion was undertaken (Figure 6.16). The *SLGE* enhancer has recently been identified as an additional gut enhancer which is comprised of 1.7 kb of sequence reported to lie ~100 kb from *Shh*. The g/120

deletion is located 96kb from *Shh* and appears to lie close to the *SLGE* enhancer and could thus capable of disrupting its function (Figure 6.16). The loss of the additional expression within the gut could therefore be as a result of the inactivation of SLGE however, what causes complete loss of pharyngeal expression remains unknown.

None of the three mouse lines resulted in a loss of expression within either the whisker buds or the skin follicles. Mice develop a coat containing four distinct types of hair in addition to several specialized hairs such as vibrissae (whiskers) and tail hair. The generation of such diversity of hair types is due to signalling pathways that drive the patterning and induction in a specific spatial and temporal manner (Duverger and Morasso 2009). *Shh* has been shown to play an important role in both embryonic and adult hair development. During embryogenesis, *Shh* is expressed in the proliferating epithelial cells at the distal tip of the developing follicle (Bitgood and McMahon 1995; Iseki et al. 1996). In mice lacking *Shh*, mature hair follicles fail to develop although hair follicle formation is initiated and the dermal condensates form, suggesting that Shh is required for the proliferation and subsequent down-growth of the follicular epithelium, and for the maturation of the dermal papillae (Chiang et al. 1996; St-Jacques et al. 1998; Karlsson et al. 1999). Treatment with a *Shh*-blocking antibody has also been shown to cause reversible alopecia, as follicles become arrested, indicating that SHH is also required for hair cycling in postnatal skin (Wang et al. 2000).

In mice, the development of vibrissae follicles starts at E12.5, and at birth, vibrissae hair shafts have already emerged. By E14.5 the wild-type whisker pad on the upper lip has been shown to consist of a vertical row of four large hair follicles alternating with five rows of varying numbers of follicles (Mill et al. 2003). Mice mutated in the *Eda/Edar* pathway (related to the tumour necrosis family (TNF)) form vibrissae at E12.5, while they do not form primary coat hair placodes at E14.5. In contrast mice mutated in the *Noggin/Lef1* pathway do not form vibrissae at E12.5 or secondary hair placodes at E16.5, but do form primary coat hair placodes at E14.5 (Botchkarev et al. 2002). Therefore, even though vibrissae are induced prior to the initiation of coat hair development, the mechanisms involved in vibrissae development seem to be more closely related to the induction of secondary coat hair

placodes than to the induction of primary coat hair placodes. Although follicle morphogenesis within the skin is completely abolished in *Shh*$^{-/-}$ mice, whisker follicles are reported to still develop in *Shh*$^{-/-}$ mice as well as in *Gli2*$^{-/-}$ mice (Mill et al. 2003).

The *MRCS1* enhancer drives Shh expression within the oral epithelium as well as the whisker buds and the developing hair follicles (Sagai et al. 2009). The three deletion lines suggest that regulation may be more complex with all three shown to retain reporter gene expression within both the whisker buds and hair follicles. Although Shh is expressed in both tissues it does not appear to have the same role, with it being pivotal for hair follicle development but not whisker barrel. Therefore, it is difficult to determine whether a single enhancer or multiple enhancers are required for driving expression within the two tissues of the hair follicles and whisker buds.

Although speculated from BAC analysis that potentially two *Shh* notochord enhancers exist within the *Shh* regulatory region 150 kb-600 kb upstream of the gene (Jeong et al. 2006) these have previously never been pinpointed. The notochord enhancer function appears to be removed between the g/2-67 and g/120 deletion suggesting these enhancers lie somewhere in the region between 100 and 530 kb upstream of *Shh*.

6.5.2 Phenotype analysis

All three deletions within the *Shh-Zrs* region were examined both as homozygotes and also with a *Shh* null allele within compound heterozygotes. From the litters examined so far no homozygous animals were found at birth, at E17.5 days or at E15.5 however they were found at E12.5 from the g/120 line (Figure 6.2). Resorbed sites were also present at E17.5 and E15.5 suggesting some degree of embryo death at these stages. The lack of embryos homozygous for these deletions could suggest that homozygous deletion of even as little as 80 kb results in lethality at later stages in development. The *Rnf32* gene resides within this region and is

deleted within all three mutants. It has also been shown to be ubiquitously expressed from early stages during development but no mouse mutants have been identified (See section 2.11). Embryonic death between the stages of E9.5-E12.5 is usually a result of placental failure, membrane abnormalities, cardiovascular abnormalities including blood vessel development and heart failure, abnormal yolk sac formation, loss of cell cycle regulation or defects in patterning (Papaioannou and Behringer 2012; Ward et al. 2012). By contrast, death between the stages of E12.5-E19.5 is often as a result of haematopoietic and neuronal abnormalities (Papaioannou and Behringer 2012; Ward, Elmore, and Foley 2012). Although the role of *Rnf32* is yet to be determined it could have a role in the haematopoietic system and thus result in mid-gestational death. The deletion of *Rnf32* within all three mouse lines could therefore be the cause of the lethality. Another ring finger gene; *Rnf213* has been shown to be involved in vasculopathies (Kobayashi et al. 2013).

The *g/2-120* line was found to result in a gut defect potentially due to proliferation defects based on the similarities between it and *Wnt5a* mutations. The g/2-67 line was shown to result in a lower jaw defect; this is likely as a result of the loss of pharyngeal and oral epithelial Shh expression and again could potentially be as a result of changes to proliferation. The g/120 line was found to exhibit a highly varied phenotype affecting the structures of the face and brain. Why there is variation with regards to phenotype remains unclear, the HPE and PPD phenotypes are highly variable with related individuals carrying the same mutation exhibiting very different phenotypes. This has been hypothesised to be as a result of genetic factors but also as a result of variations in SHH protein levels within individuals. It is believed that there exists a threshold for SHH protein levels above which no phenotype is observed. However if protein levels are found to be below the threshold severe malformation can occur. Inheritance of HPE and PPD in humans is complex as haploinsufficiency is found at the *SHH* locus, however, variations have also been identified in mice. Recently it has also been shown that attenuated expression produced by a *Zrs* 3′ end deletion within mice revealed variability inherent in regulating digit number. The variation in the limb phenotype ranged from a *Shh* null phenotype in the fore limbs to two to four digits with deletions and fusions of tarsal

and metatarsals in the hind limbs and has been attributed to protein levels and temporal extent of exposure (Lettice et al. 2014). Temporal changes in when Shh expression is lost could also go some way to explain the variation seen within the *g/120* line. Shh is known to be expressed within several signalling centres within the face with expression from the forebrain being important at early stages for forebrain development and at later stages for development of the upper jaw and nose structures. It has previously been shown that the timing of Shh loss within the forebrain can affect whether a forebrain or facial defect is observed (Marcucio et al., 2005, Chong et al., 2012).

Genetic rescue could also play a role in the observed varied phenotype. As only a low level of Shh protein is required to reach a normal development threshold it is possible that expression from yet unidentified redundant/shadow enhancers or due to rescue from other sources of Hh could result in varying the phenotype. This could suggest why, for example, the mildly affected g/120 embryos appeared to have rescued the gut phenotype that is present in the smaller deletions. However expression analysis within both compound heterozygous and g/120 homozygote embryos suggests that a different mechanism could be at play. Staining of these embryos to examine reporter gene expression shows that within those with a severe phenotype reporter gene expression is lost from all tissues except those of the limb. This could suggest that the tissues such as the floorplate, mid-brain and gut have not formed. The notochord source of Shh extends into the floorplate and brain where it is required for ventrally patterning a number of cell types within the hindbrain and midbrain. Loss of this source of Shh results in collapse of the brain structures (As previously described Section 2.3). The g/120 deletion has been shown to result in loss of reporter gene expression within the notochord suggesting one or several enhancers are present within the deleted region which could drive expression here. It is possible therefore that the severe phenotypes are as a result of loss of both forebrain and notochord expression whereas in the less affected animals some or both of these tissues have been rescued.

The alternative explanation is that the g/120 deletion disturbs the region in a way that even enhancers which have not been deleted are affected. This seems unlikely; expression analysis within heterozygous animals suggests the remaining enhancers function normally both temporally and spatially (Figure 6.10), additionally within the homozygous and compound heterozygous animals the *Zrs* is still found to function (Figure 6.10). Previous reports within the literature suggest some developmental loci function as archipelagos or holoenhancers where each of the component parts being important for the overall expression pattern (Discussed section 2.15.1). In such a scenario loss of one component part can result in an uncoupling of the entire domain. Although redundant/shadow enhancers have been identified at the *Shh* locus, no evidence has so far suggested that it behaves as a holoenhancer or archipelago, instead it is proposed that each of the individual enhancers will function regardless of the presence of the remaining enhancers.

6.5.3 The effect of multiple promoters on enhancer activity

Mouse lines were also generated in order to examine the effect of multiple ectopic promoters on enhancer function. Mice homozygous for both the SBLac2-120 and SBLac2-67 lines appeared developmentally normal and were capable of recapitulating a LacZ expression pattern similar to the heterozygote (Figure 6.14). These data, therefore, suggest that it could indeed be possible for an enhancer to contact two promoters within a single tissue, as shown by the fact that the enhancers within SBLac2-120 and 2-67 homozygote embryos are capable of contacting LacZ as well as *Shh* (as suggested by the lack of phenotype). However the phenotype and changes in expression pattern exhibited within SBLac120 homozygotes suggests this is not always the case. The overall expression pattern within SBLac120 homozygote embryos was found to be less intense than that seen within the heterozygote despite carrying two copies of *LacZ,* suggesting that although the enhancer is capable of driving both the promoters within the same tissue, it is at a reduced amount compared to times when there is less promoter competition (Figure 6.14).

Both the forebrain and facial enhancers appear to be more susceptible to the interference of the extra promoter within SBLac120 homozygotes as compared to the enhancers such as the *Zrs* as well as those within and close to the *Shh* gene. This could suggest that these enhancers are simply less robust than the others within the region and serve to confirm the variation between the enhancers. Alternatively this could be as a result of directionality; the two *LacZ* promoters are located 100 kb from *Shh* within this line and thus lie between the promoter and both the forebrain and facial enhancers. The floor plate and midbrain enhancers sit after these additional promoters and therefore would not necessarily be 'blocked' from reaching the *Shh* promoter by them. The ZRS is located at a large distance from *Shh* and is proposed to interact via chromosomal looping meaning it would be possible bypass the extra promoters within the region.

The fact that such an extreme phenotype is found within the SBlac120 homozygotes whereas apparently no phenotype or change in expression pattern is exhibited within either the SBLac 2-120 or SBLac 2-67 lines suggests that the position of these extra promoters is important in how they affect the region (Figure 6.14A). Both of these insertion sites are downstream of the forebrain enhancers; however, they both sit between the pharyngeal enhancer MFCS4 and *Shh* which is also likely to be involved within the craniofacial phenotype shown within the SBLac120 homozygotes considering expression within the pharyngeal structures was shown to be disrupted. This suggests that there are sites within the *Shh* region which are more susceptible to being disrupted by multiple promoters. The phenotypes seen within the g/120 mice must therefore be taken with caution, not only could these be arising as result of the deletion but the multiple promoters present within this site could also complicate matters and potentially are another factor which could be accountable for the variation in phenotype found within this line.

7. Discussion

7.1 Introduction

The *Shh* genomic region encompasses a complicated regulatory locus comprising a series of regulatory elements which serve to drive expression of the gene within specific tissues of the central nervous system, gut and posterior of the limb bud (Epstein, McMahon, and Joyner 1999; Lettice et al. 2003; Sagai et al. 2004; Jeong et al. 2006; Sagai et al. 2009). These enhancers lie within the gene itself, within a gene desert upstream of the gene and also within the introns of neighbouring genes, spanning a distance of nearly 1 Mb. A great deal of work has been performed on developmental loci including the *Shh* domain in recent years with an aim to understand how these developmental regions function.

7.1.1 Targeting of an LHED vector into the *Shh* region

Utilisation of the LHED system (Kokubu et al. 2009) has enabled us to study the *Shh* region in depth. Firstly, three separate LHED vectors were designed and cloned, one within the 5´end of *Lmbr1* and the other two within the 3´end. Of these, the 5´ end and one of the 3´ end vectors were successfully targeted into E14 ES cells. The remaining vector could not be successfully targeted into the correct genomic region. When designing the homology arms for targeting, two specifications were taken into account firstly the homology length and secondly how much unique DNA is included within this homology as both are important to allow the cellular DNA repair mechanisms to align a targeting vector with its corresponding region of homology and result in recombination into the chromosome. For recombination to occur within a cell, around 2 kb of sequence homology is required (Melton 2002). However, 6-14 kb of homology is typical for targeting constructs. Although the 3' targeting vector which failed to recombine was 4.6 kb and thus should have had a long enough homology arm to insert within the genome it was not possible to detect this. The

actual targeted event takes place in only a small percentage of cells (Hall et al. 2009). During a stem cell experiment, only about 10^{-2} to 10^{-3} of the DNA integrations are homologous recombination events (Melton 2002). Thus it is possible that simply not enough clones were analysed to detect positive insertion of this vector.

Vectors with a double stand break within the homologous counterpart of the chromosomal sequences have been termed insertion vectors and usually recombine through a single reciprocal double stand break repair pathway. A vector that targets via this pathway will integrate the entire vector to form a duplication of the homology arm. The utilisation of primers homologous to sequence within this double strand break was found to be the only reliable way of determining correct targeting of the vector into the genomic locus. It is thus essential for this unique piece of DNA to be repaired in order for the full homology arm to become inserted within the genome (See Figure 4.4).

7.1.2 Remobilisation of the LHED vector

The process of remobilising the LHED vector *in vitro* was optimised. It was found that the SB100x transposase (Mates et al. 2009) was much more efficient than the SB11 transposase within the original study (Kokubu et al. 2009). A chemical transfection method rather than electroporation was also found to be more reliable for treating LHED-carrying cells with the transposase vector. Finally, it was discovered that a higher level of puromycin was required in order for selection to occur.

Within the original LHED study (Kokubu et al. 2009) a splinkerette-type PCR was used for screening ES clones in order to establish where reinsertions had occurred. However a somewhat simpler method to establish Sleeping Beauty reinsertion within mouse lines was more recently detailed by (Ruf et al. 2011). This involved an asymmetric nested PCR and sequencing and was, therefore, employed in order to analyse the LHED clones generated from both targeting vectors. This system appeared to be highly reliable; altogether it was possible to map sequence for nearly

90% of clones analysed and indeed this number could potentially be increased by repeating the PCR and sequencing (See appendix Table 8.1). Approximately 40% of these reinsertions were found to occur within 1Mb of the original insertion site (Table 4.2). This provided a good distribution of re-insertions across the region (Figure 4.8). One limitation of the system was that the orientation in which the transposon reinserted was important for whether the *LoxP* sites could be utilised for generating deletions and maintain the LacZ reporter gene. Although there was a roughly 50:50 ratio for orientation of reinsertion, there was, in fact, a somewhat limited number of insertions which proved useful for generating deletions. In order to map the region more thoroughly it may be required to generate more insertion sites, either by targeting another vector perhaps in the middle of the region, which could be done with relative ease now the system is completely optimised. Alternatively, one of the existing insertion sites could be re-mobilised by additional treatment with the transposase, the only limitation of this being that it will not be possible to utilise the Puromycin resistance cassette to determine those clones in which mobilisation has occurred.

7.2 Use of the LHED system *in vitro*

One of the benefits of using the LHED system over other Sleeping Beauty systems such as Gromit (Ruf et al. 2011) was the ease of using it within ES cells rather than mice. Although targeting the LHED system into the ES cells and subsequent mobilisation of the transposon was somewhat laborious, the fact that it allowed selection of appropriate clones was helpful to the further study. This meant the ES cell lines could be split into two categories, those that were beneficial for generating tetraploid embryos (fairly quickly) to examine reporter gene expression and thus map the region, and those which were to be used to generate mouse lines to look at the effect of deletion on phenotype (more time consuming).

7.3 Mapping the *Shh* region using LHED

A series of LHED insertions were used to generate tetraploid embryos in order to look at how reporter gene expression varies across the *Shh* region, this was done with an aim of looking at reporter gene expression within the gene desert, within neighbouring genes as well as within the boundary regions flanking the locus.
Reporter genes were inserted into four positions throughout the 729 kb gene desert, analysis of which suggest that the desert is in an open conformation over its length and each reporter is receptive to all the known enhancer activities. Each of the inserts recapitulated spatial patterns that reflected *Shh* expression upon examination of the whole embryo and the dissected gut. This suggests that the gene desert flanking *Shh* is broadly open to transcriptional activity. Additionally, the spatial pattern of expression was independent of the position of the reporters within the desert suggesting that there is no segmentation of the regulatory terrain into open and closed chromatin and that the chromatin within the gene desert is functionally indistinguishable.

Although the global expression pattern within each of the desert inserts suggests a lack of variation, the activity levels are not equivalent across the domain. The most distal enhancer, the Zrs limb -specific enhancer lying in the *Lmbr1* gene showed a trend in which the reporter closest to the enhancer was expressed at a high level with a decrease in expression toward the middle of the domain and an increase with the reporters near the *Shh* gene. This increase in pLHED-derived reporter activity nearer the gene was also reported for the *Pax*9 gene (Kokubu et al. 2009). Alternatively, the *MACS1* gut enhancer is found within the *Rnf32* gene (Sagai et al. 2009) and expression driven by *MACS1* in the laryngotracheal tube suggests a mechanism more focused in the direction of the gene working less efficiently closer to the enhancer itself. Finally, the *SBE1* enhancer which lies within the *Shh* gene was found capable of working across the entire gene desert suggesting its proximity to the target promoter is of no consequence for its activity.

Analysis of the *Fgf8* locus has suggested that interactions between multiple enhancers may lead the region to adopt a defined structural conformation that could favour regulatory interactions between enhancers and specific positions of the locus

instead of with specific promoter sequence (Marinic et al. 2013) (See section 2.15.1). Although it is possible to disassemble these regulators into autonomous modules, the whole region is speculated to work as an integrated regulatory assembly, whereby overall activity cannot simply be determined by the addition of its basal components. In this model, dynamic interactions between the different active elements, involving eventually additional interspersed structural elements, determines where the regulatory influences are exerted; hence, the output of the entire region would reside in the relative distribution of genes and enhancers along the locus. Within the 'holoenhancer' system the reporter gene frequently showed only a subset of the expression of the endogenous gene, suggesting that although all domains could be captured by an endogenous promoter, they were not (Marinic et al. 2013). From this it was suggested that regulatory elements that control gene expression may have distinct ranges of action, thereby defining different regulatory landscapes and ultimately resulting in differential genes expression at different positions within a locus, as a result of the functional consequence of chromatin and conformational structures as well as the spatial range of enhancers and of their extensive interactions (Marinic et al. 2013). What is reported for holoenhancers seems to contrast with what is seen at the *Shh* locus, indeed there does not appear to be any correlation between the position of the reporter and activity, with enhancers seemingly acting regardless of location and not always focussed towards the gene in question. Additionally, none of the inserts examined both within genes and also within the gene desert showed only a subset of the gene expression pattern, instead they were found to recapitulate the pattern (although with tailing off at the boundaries) suggesting that if the promoter of the reporter gene was in contact with the enhancer(s) they acted upon it.

Further, Gromit (Ruf et al. 2011) also suggests that within a regulatory domain, the expression level detected by the reporter can vary quantitatively and reveal positions apparently refractory to activation known as cold-spots. The purpose of these cold-spots is speculated to be a way to shield genes from the influence of surrounding enhancers (Marinic et al. 2013), adding to other mechanisms of specificity such as core promoter sequences (Ohtsuki et al. 1998). Again a different

scenario is found within the *Shh* locus. The *Shh* regulatory domain contains neighbouring genes including the ubiquitously expressed *Rnf32*. Although the gene body of *Rnf32* appeared to be open to enhancer activity the promoter was resistant. A similar situation was also found downstream of *Shh* whereby the ubiquitously expressed *Rbm33* gene located within the *Cnpy1/En2* was open to enhancer activity although the promoter appeared refractory. This therefore suggests a common mechanism within large regulatory domains may be that ubiquitously or widely expressed genes are not sequestered by insulators or boundaries but are able to occupy domains within the range of enhancer activity by containing promoters that are simply resistant to remote regulatory activity. Overall the general conclusion from analysis of the *Shh* region suggested that it is not acting as a holoenhancer.

The boundaries surrounding the *Shh* region were also examined and defined by looking at insertions both within and upstream of the *Lmbr1* gene as well as downstream of *Shh*. At the 5´ end of the region a more gradual decline in enhancer activity was found where individual enhancers were lost in increments. The 3´ end of the region seemed more definitive whereby loss of all enhancer activity was found fairly closely downstream of *Shh*. A similar situation was found at either end of the region whereby as soon as *Shh* expression was lost a new expression domain was found. Downstream of *Shh* this involved *Cnpy1/En2* whereas upstream of *Lmbr1* an expression pattern likely to involve *Mnx1* was identified. There was little overlap between the domains however it was possible that the *Zrs* was still functioning upstream of *Lmbr1*, long after the other *Shh* enhancer activity was lost. The region appears to comply with the Topological Activity Domain (TAD) previously suggested (Smallwood and Ren 2013) (Figure 5.8). Within this model the *Lmbr1* end of the TAD appears quite ‘messy’ which also both fits with the data found and could explain the overlap in expression between the *Shh* and *Mnx1* domain. Recent work by (Symmons et al. 2014) has shown that despite the overall inclusion of regulatory domains into TADs, the positions of the relative transitions were not always exactly superimposed, similarly to chromatin domains, which do not exactly correlate with TADs (Hou et al. 2012). These differences have been suggested to partially arise from the low resolution of Hi-C, which can locate topological transitions with only

limited precision (20 kb). Furthermore, topological transitions may not necessarily constitute absolute barriers, but act more like dampers (Andrey et al. 2013).

Many known insulator or barrier elements are bound by the zinc-finger containing protein CTCF. Within the *Shh* and *Lmbr1* region a number of such sites were found, and notably, these lie not in the gene desert but are enriched at either end of the regulatory domain. However there is no direct relationship between the position of the CTCF sites and the loss of enhancer activity. The conditional loss of CTCF within the forelimbs does not suggest it acts as an insulator at least in developing limb mesenchyme (Soshnikova et al. 2010). A similar global analysis has been performed by (Symmons et al. 2014) which also suggested that, by and large, it appeared difficult to match the distribution of the CTCF/cohesin sites with the distribution of regulatory activities, highlighting that, within TADs, the binding of cohesin/CTCF, defined by chromatin-immunoprecipitation, may not be a sufficient indicator or regulator of regulatory boundaries.

7.4 The effect of deletions on reporter activity

The LHED system was also used to generate deletions within the *Shh* region (*g/2-120*, *g/2-67* and *g/120*); the effect of these was then determined by examining reporter gene expression. Embryos were examined between stages E9.5 and E14.5, none of the deletions reported however showed any temporal changes in expression at these stages, only spatial.

Deletion of enhancers with known expression patterns appeared to result in a fairly consistent loss of reporter gene expression within those tissues involved. For example loss of *MACS1* resulted in a down-regulation of expression within the gut, loss of the enhancer *SBE2-SBE4* resulted in loss of expression within the forebrain tissues, whereas loss of *MFCS4* appeared to result in at least changes with respect to expression within the pharyngeal arches and tongue. No aberrant or spurious expression patterns were identified following deletion of DNA from the locus suggesting that removal of the various regions did not result in gain of any enhancer

functions. In a similar respect no changes in expression pattern of any remaining enhancers were identified.

Reporter gene expression following deletion suggested the possibility of shadow enhancers within the region, which are expected to help foster robustness in gene expression. Within both the smaller deletions, although gut expression is reduced following deletion of the *MACS1* enhancer, is it not completely lost (Figure 6.3). A second gut enhancer *SLGE* is also present within the region (Tsukiji, Amano, and Shiroishi 2013) and although this is not necessarily deleted in full within the large deletion, it is disrupted and thus could be accountable for the complete loss of gut expression (Figure 6.9). A similar situation to the gut was found for the pharynx whereby expression was shown to be reduced following the g/2-120 and g/2-67 deletions before being completely lost with the g/120 deletion (Figure 6.4). *MFCS4* has been shown to drive expression within the pharyngeal tissues and no additional enhancers have so far been identified. These data however suggest an unknown shadow enhancer potentially exists, which acts alongside *MFCS4*, or alternatively. perhaps *SLGE* has additional function beyond just the gut.

The *MRCS1* enhancer was previously reported to drive Shh expression within tissues of the oral epithelium, as well as hair follicles and whisker buds (Sagai et al. 2009). However, this was found not to be the case, the two deletions in which this enhancer was removed were still found to exhibit reporter gene expression within both the hair follicles and whisker buds (Figure 6.6/6.9). *MRCS1* therefore does not adhere to the expression pattern previously reported. *In vivo* reporter experiments often rely on "minimal" enhancer fragments which do not always perfectly replicate the precise spatial boundaries of expression of the native gene in the endogenous context, thus it is entirely possible that the expression pattern reported for *MRCS1* does not necessarily reflect the true expression pattern of the enhancer. Alternatively it could suggest the presence of one or more shadow enhancers are capable of driving expression within these tissues. Within the *Shh* null mouse hair follicles do not form although whisker buds do (Mill et al. 2003). This suggested the two developmental pathways for these tissues are independent of each other. It also suggests at least in the case of whisker buds and Shh expression is not essential for development.

SBE1 is responsible for the maintenance of *Shh* expression in the midline of the midbrain and caudal diencephalon from E10.5 onwards; is believed to act in concert with an additional enhancer which controls *Shh* transcription at earlier time points as well as driving expression within the Zli (Jeong et al. 2011). Within all three deletions mid-brain expression is present from E9.5 onwards and expression is shown within the zli suggesting that this additional enhancer is not present within the region which has been removed. The presence of a notochord enhancer was also confirmed in this experiment via the loss of notochord expression between the middle and largest deletion. The presence of two notochord enhancers has previously been postulated from enhancer trap analysis (Jeong et al. 2006), these data therefore serves to both confirm the presence of these and also narrow down the region in which they lie.

7.5 The effect of deletions on phenotype

Three deletions were generated to examine phenotype, all of which were shown to be lethal before birth as both compound heterozygotes with a *Shh* null chromosome and also as homozygotes. Within the homozygotes this lethality could be as a result of the presence of *Rnf32* as these animals appeared to die at earlier stages than the compound heterozygotes.

The g/2-120 deletion which was the smallest deletion was found to result in a gut defect as a result of loss of some gut enhancer function during development. The guts were shorter and the stomachs were malformed, however the phenotype was not found to be similar to the *Shh*-null gut defects. These embryos were found to have residual LacZ expression within the guts suggesting enhancer function (potentially *SLGE*) within this tissue had been reduced but not lost. Loss of enhancer function within the gut was not actually achieved until the final deletion g/120 which was found to result in a gut defect similar to the *Shh*-null. The lungs of g/2-120 animals appeared structurally normal as compared to *Shh*-null animals which are completely

devoid of lungs. The lack of a lung phenotype within the g/2-120 deletions seemed strange considering that although there was residual reporter expression within the lungs of these animals, a large decrease in enhancer activity comparable to the rest of the gut had occurred. This lack of phenotype could therefore suggest that the lungs are less sensitive to reductions of Shh protein levels or are rescued more easily by other sources of Hh. Within some of the g/120 homozygotes a surprising phenomenon occurred in that reporter gene expression was found to be present within the guts (despite not being present within the heterozygotes) (Figure 6.10). Additionally, older g/120$^{+/-}$; Shh$^{-/+}$ embryos exhibiting milder facial defects did not appear to have the same gut defect as seen with the two smaller deletions suggesting the phenotype had somehow been rescued (Figure 6.11). This rescue of phenotype/expression again suggests the presence of shadow enhancers present within the region which are activated upon the loss of *MACS1*. The SBLac120 insertion occurred within the enhancer *SLGE* which potentially functions as a shadow enhancer to *MACS1*. Following the deletion, loss of enhancer activity was found to occur within the guts of g/120 heterozygous mice suggesting this enhancer was no longer functioning, however, some of the SLGE sequence remains, which could be functioning to turn on *SHH* expression thus rescuing the phenotype. However, why this rescue of expression pattern and phenotype within the smaller deletions does not occur is unclear.

Both the g/2-67 and g/120 deletions resulted in animals with craniofacial defects. Within *g/2-67$^{+/-}$; Shh$^{-/+}$* and at the milder end of the phenotypic scale of *g/120$^{+/-}$;-Shh$^{-/+}$* mice this resulted in a phenotype comprising a truncated lower jaw and an absent tongue, these mice were also found to exhibit a pointed face, attributes reminiscent of failure of both lower jaw development (derived from pharyngeal structures) and also late forebrain expression. However, the brains of both lines exhibiting this facial defect were found retain their structures. Deletion of the pharyngeal enhancer both within a previously published study (Sagai et al. 2009) and also within the g/2-120 mutant suggests that a jaw phenotype is not found unless both this pharyngeal enhancer as well as the oral epithelial enhancer is removed. This

likely suggests that the tissues in which both of these enhancers are being expressed are affected by the loss of Shh, resulting in the facial phenotype.

The *g/120* mouse line was found to result in a varied phenotype. Genetic rescue of tissues in which reporter gene expression has been abrogated could explain the phenotypic variance as a result of Hh expression from other tissues. Ihh expressed from the gut endoderm has previously been suggested to be capable of rescuing CNS development including defects in the telencephalon (Zhang et al. 2001). Temporal effects could also be the cause of the varied phenotype seen within the *g/120* line. Previously, studies have shown that early removal of SHH from the forebrain results in severe forebrain defects; whereas, later removal is restricted to malformations of the upper jaw (Marcucio et al. 2005; Chong et al. 2012). Reporter gene analysis suggests that within the mutant forebrain expression is lost from E9.5 days onwards (Figure 6.9) and analysis of $g/120^{+/-}; Shh^{-/+}$ compound heterozygotes suggests that this loss of forebrain expression is found in E12.5 embryos exhibiting moderate phenotypes (Figure 6.10). Whereas within those with a more extreme phenotype reporter genes expression is not only lost within the forebrain but also within the remaining structures of the CNS. The notochord source of Shh is important for floor plate as well as hind brain and mid brain development (Kaellen 1965; Britto, Tannahill, and Keynes 2002) and has also been shown to be lost within g/120 mutant mice (Figure 6.9). This could suggest the varied phenotype found within the g/120 line could be as a result of the compounded effects of the loss of forebrain and notochord Shh sources.

Shh protein levels are also likely to be important within $g/120^{+/-}; Shh^{-/+}$ mutants. Variations in limb phenotype occur following deletion of part of the *Zrs* as a result of temporal changes or as a result of changes in Shh protein levels (Lettice et al. 2014). Within individuals with HPE; phenotypic variation has been attributed to threshold levels of Shh protein (Jeong et al. 2011). Although enhancer function has been shown to be lost/reduced following the deletions within the three lines Shh protein could still be present. Variations in phenotype could therefore be as a result variations in protein levels and whether or not Shh threshold is met.

Despite deleting the large gene desert between *Shh* and the *Zrs*, no alterations in either *Zrs*-driven reporter gene expression or limb phenotype were observed, suggesting the distance at which the *Zrs* operates at is not important for function. As previously discussed, recent studies at other developmental loci suggest that regulatory domains act together in order to drive expression of their target gene (See section 2.15.1). Marinic et al., (2012) suggest that in intact endogenous loci regulatory control is exerted by coordinated and not individual action, of enhancers suggesting integration is orchestrated at the level of the regulatory domain, not at endogenous gene promoters. Genetic dissection of the 800 kb large genomic intervals centromeric to HoxD13 indicated that various elements contribute in a partially redundant fashion to the transcriptional activation of HoxD genes (Montavon and Duboule 2013).

It is unknown if either of these models are applicable to the *Shh* domain, the information generated from mapping the region suggests that the regulatory domain is primarily composed of regulators that contribute to the spatiotemporal expression pattern as a summation of the individual enhancer activities. Deletion of these various elements also confirms this whereby each of the individual enhancers deleted resulted in a loss of the spatio-temporal expression from such an enhancer without resulting in any changes to any remaining enhancer activity or the overall output of the domain. The differences between the Shh domain and other regulatory loci such as *Fgf8* and *Hoxd* could therefore suggest that the coordinated gene expression controlled by long distance enhancers falls into a number of different but possibly, overlapping mechanistic classes.

7.6 The effect of multiple promoters on enhancer activity

A final strand of this work involved examining whether the presence of multiple promoters affected reporter gene expression and also phenotype. This was done with an aim of examining whether enhancers were capable of contacting multiple promoters within a single tissue. This is believed to occur at loci as ‘promoter competition’ is speculated to provide a way to filter out enhancer activity from target

promoters as described at the HoxD locus (Spitz, Gonzalez, and Duboule 2003). SBlac120 mice were crossed in order to generate homozygous individuals thus containing a *LacZ* reporter gene and corresponding *Hsp68* promoter on both chromosomes. Embryos were shown to have disruption of forebrain and pharyngeal reporter gene expression resulting in a craniofacial defect (Figure 6.13/6.14). Homozygous embryos generated for the two other inserts SBLac 2-120 and SBLac 2-67 had no changes to reporter gene expression or phenotype as compared to heterozygous controls (Figure 6.13). The presence of comparable LacZ expression and phenotype within SBLac 2-120 and SBLac 2-67 heterozygotes and homozygotes suggests that it is possible for an enhancer to contact both the endogenous and an additional promoter within a single tissue. However, as the overall expression pattern within SBLac 120 homozygous embryos was found to be less intense than that seen within the heterozygotes despite carrying two copies of *LacZ* than some degree of promoter competition occurs at least at some sites within the region resulting in a phenotype.

A phenotype was found within SBLac120 homozygous mice but not within either of the other homozygous insertions suggests that some sites within the *Shh* region are more susceptible to being disrupted by multiple promoters than others. This is maybe a result of complicated chromosomal dynamics within this locus which are yet to be elucidated.

7.7 Perspectives and further work

By utilising the LHED system it has been possible to generate a series of SBLac insertions throughout the *Shh* regulatory region. The system has also been optimised which provides a further possibilities of exploiting the system in order to answer a great deal of possible questions. A number of the insertions generated were made use of in order to map regulatory activity across the gene desert, and at the boundaries of the region by employing a tetraploid complementation assay. Additionally three mouse lines have been generated in order to examine loss of function effects within the region. The mutations generated within these mice were shown to reflect defects of gut and craniofacial development; although numbers and

generation remain low therefore further work will be required to further verify this and examine phenotype on additional backgrounds. A great deal of variation was found with regards the largest deletion (g/120) although again a limited number of embryos were studied. A number of possible hypotheses have been drawn with regards to the cause of this variation including differences in temporal expression, variations in protein levels and rescue of expression/phenotype however potentially a multitude of these reasons could be in play. Finally analysis was performed on homozygote SBLac embryos, from this it was concluded that certain sites within the region are more susceptible to the influence of multiple promoters. This suggested another potential caveat which could play a role in the phenotypic variation seen within the g/120 line. Again however embryo numbers will require to be boosted in order to really study this in depth.

With regards future work further backcrossing of all the lines is required in case variation is background-dependent. Further characterisation of the phenotypes generated must also be undertaken using further antibody or *in situ* markers. Additionally elucidation of how Shh still contributes to cells within the deletion lines would also be useful. GFP/Cre *Shh* 'knock-in' mouse lines exist which could be crossed with the deletion lines in order to recombine both the deletion and the GFP/Cre *Shh* alleles onto the same chromosome and thus produce double marked LacZ/GFP cells.

As the presence of multiple LacZ promoters at certain locations has been shown to affect phenotype it would also be beneficial to test the effect of LacZ removal on the lines. This currently cannot be done in the current arrangement therefore re-engineering would have to be undertaken in which case it may be beneficial not to delete within the middle of the SLGE enhancer.

Analysis within the region has also suggested the presence of a number of additional enhancers or shadow enhancers for example those which control notochord expression; fine mapping of the region by generating smaller deletions would potentially allow narrowing down of their location. Enhancer marks could also be used to pinpoint enhancer sequences and reporter transgenics utilised to test enhancer function. The effect the deletions have on the *Shh* topological domains would also be

of interest, this could be investigated by employing 5C. Finally, the CRISPR system (Sander and Joung 2014) is currently being used within the lab with success and therefore it may be beneficial to combine both this and the LHED system for further analysis of the *Shh* region.

8. Appendix

8.1 Developing a FISH/ES cell system to establish the chromosomal dynamics at the *Shh* locus

Although the LHED system was initially employed to be used to generate tetraploid embryos and mouse lines, the use of ES cells also meant it was possible to do some analysis within the cells themselves. The *Shh* locus has been shown to interact with the ZRS via chromosomal looping within cells of the ZPA (Amano et al. 2009). It is possible that the composition of the *Shh* regulatory landscape is important for this interaction.

8.1.1 Inducing Shh expression within ES cells

Wild-type ES cells do not express Shh. Therefore, in order to induce Shh expression, cells were treated with 1 μM retinoic acid over a period of 72 hours. Retinoic acid causes the cells to differentiate down a neuronal pathway. At 24 hours intervals RNA was extracted from these cells as well as a non-treated control and used to produce cDNA. The resulting cDNA was then analysed by RT-PCR using primers for *Shh* as well as the differentiation marker *Oct4*. Expression of *Shh* was detected following 48 hour treatment with retinoic acid (R.A). Whereas, *Oct4* expression was shown to gradually decrease over the timescale indicating the cells were differentiating (Figure 8.1A).

8.1.2 What are the chromosomal dynamics at the *Shh* locus?

FISH analysis was undertaken in order to examine the configuration of the *Shh* locus both in wild-type and Shh expressing ES cells. Metaphase nuclei were prepared from both control cells and those which had been treated with R.A for 48 hours. FISH was then performed using probes against *Shh*, the *Zrs*, *Rnf32*, *SBE4* and *Dpp6*. Within wild-type ES cells *Shh* and the *Zrs* were found to have a much smaller interprobe distance than *Shh* and *Dpp6* (Figure 8.1B [i] [ii]). *Rnf32* and *SBE4* were

also found to be within close proximity to *Shh*, thus suggesting the whole locus is held in a compact configuration rather than a 'looping-out' event occurring between the *Zrs* and *Shh* (Figure 8.1B [iii] [iv]). In comparison, within Shh expressing cells, the interprobe distance between *Shh* and the ZRS was found to be substantially smaller than that between the same probes in wild-type cells (Figure 9.1B [v]). This therefore suggests that expression of Shh causes an even tighter compaction of the *Shh* locus. The Zrs is not believed to be the target of retinoic acid as it is not associated with the neuronal pathway. A likely target is one of the neural *Shh* enhancers.

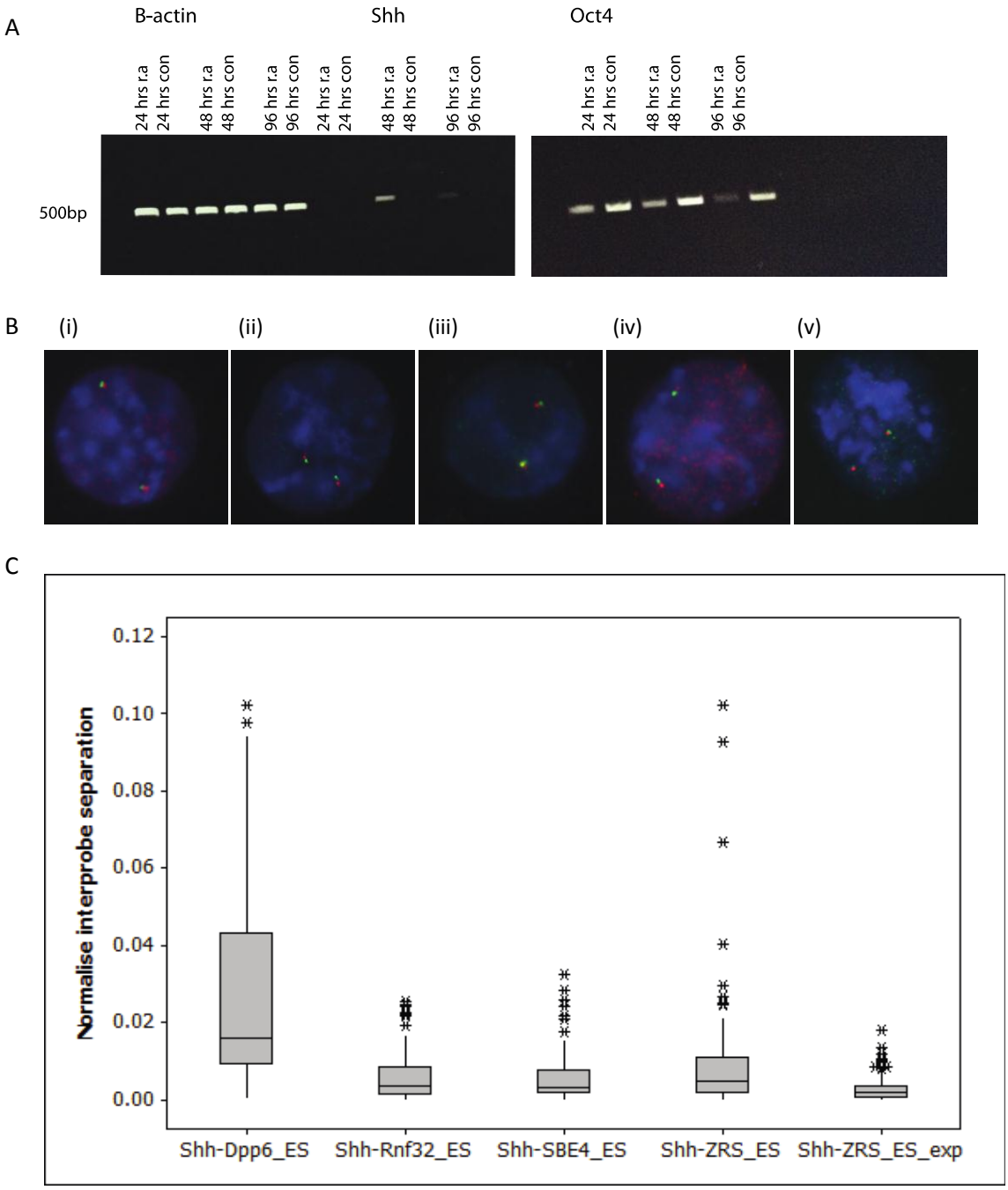


Figure 8.1: Chromosomal dynamics at the *Shh* regulatory locus

A. RT-PCR expression profile of *B-actin*, *Shh* and *Oct4* within Wt and R.A treated ES cells. *B-actin* levels shown to be consistent in all cell types. *Shh* is shown to turn on in ES cells treated for 48 hours with R.A and a small amount is still found at 96 hours. *Oct4* is shown to slowly turn off following retinoic acid treatment.
B. FISH analysis of ES cells with probes for *Shh* (green) and *Zrs* (red) in wt cells (i) and Retinoic acid treated cells (v). FISH also shown for *Shh* (green) and *Dpp6* (red) (ii) and *Rnf32* (red) (iii) and SBE4 (red) (iv).
C. Box-plot of average interprobe distances over an average of 100 cells. *Shh-ZRS* is shown to be much more compact than *Shh-Dpp6*, this is shown to reduce further following R.A treatment (*shh-zrs exp*).

8.2 Generating deletions in vitro to examine the chromosomal dynamics at the *Shh* locus

In order to examine whether there is anything special about the *Shh* locus which affects its dynamics, four of the LHED insertions were taken and used to generate deletions within the cell lines. These included a series of three insertions between the Z*rs* and *Shh* (2-120, 2-67 and 120), located approximately 100 kb, 500 kb and 700 kb from *Shh*. A further insertion 100 kb upstream of *Lmbr1* was also chosen (3-119) as it would allow deletion of the transcriptional start site of *Lmbr1* and potentially any boundary elements in the region.

$1x10^7$ cells from each line was taken and electroporated using 100 µg of a *Cre recombinase* expressing vector. $1x10^5$ cells were then plated out and left to grow on antibiotic free media for approximately 7 days. 100 clones were picked per transfection and PCR analysis employed to determine if the deletion had occurred using primers designed either side of the *LoxP* sites which would only provide a PCR band if the deletion had indeed occurred. Approximately 10% (10/100 picked) of clones for each transfection were found to carry the appropriate deletion (Figure 8.2).

Positive deletion cell lines were grown up and used to produce metaphase nuclei for FISH analysis.

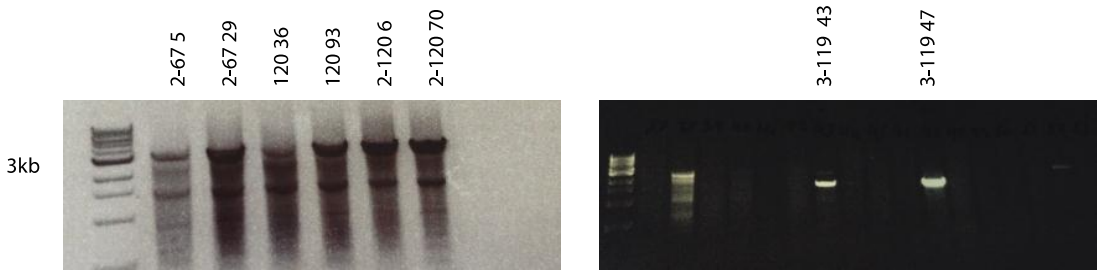


Figure 8.2*: in vivo* LHED deletions within the *Shh* region

Long range PCR analysis of LHED targeted lines treated with *cre recombinase*. Primers were designed to cross the deleted region generating bands at 3.5 kb.

8.3 Generation of deletion lines upstream of the ZRS

The predicted protein product of the *Lmbr1* gene is a novel multipass transmembrane protein that does not fall into any known functional class but has been highly conserved in different organisms (Clark et al. 2001). Its structure suggests that it may encode a membrane anchoring protein, adhesion molecule, transporter, or cell surface receptor (Clark et al. 2001). *Lmbr1* which is located within the *Shh* region and contains the *Zrs* within one of its introns appears to be ubiquitously expressed during development. To test the requirement for the *Lmbr1* gene during development, a mutation in the *Lmbr1* gene has been created that deletes the first known exon that contains the predicted site of translational initiation for both known *Lmbr1* products (Clark et al. 2001). Mice carrying this allele show greatly reduced expression of *Lmbr1* transcripts but still express low levels of novel transcripts. Mice heterozygous or homozygous for the $Lmbr1^{ATG}$ mutations are viable and fertile and do not show typical *Zrs* related limb defects. Instead, homozygous mutant mice show a very low incidence of limb defects, including oligodactyly, reduction in length or number of phalanges, and soft tissue or bony syndactyly. The placement of enhancers within functional neighbouring genes is not a novel concept and a similar phenomenon has been found in several developmental loci including those of *Pax6* and *Gremlin* (Kleinjan and van Heyningen 1998; Zuniga et al. 2004). Indeed at the *Shh* locus, the gut enhancer *MACS1* is also found within the *Rnf32* gene as well as several brain and floor plate enhancers within the *Shh* gene itself (Epstein, McMahon, and Joyner 1999; Jeong et al. 2006; Sagai et al. 2009). Whether there is a functional reason for these enhancers being within expressed genes remains unclear. One possible explanation could be that the enhancers benefit from being within a region of open chromatin. In order to test this theory a series of mouse lines were generated in which the 5´ end of the *Lmbr1* gene was disrupted.

Two LHED insertions were used to generate mouse lines which would allow disruption of the *Shh* regulatory locus upstream of the *Zrs*. These insertions were firstly, SBLac3-119 located approximately 100 kb away from *Lmbr1* which could be used to generate a large deletion removing the 5' end of *Lmbr1* as well as the next gene *Nom1*. A second insertion SBLac3-21 was also decided upon, this is located

approximately 25 kb upstream of *Lmbr1* and within the orientation which would allow the 5' end of *Lmbr1* to be inverted without deleting anything of importance.

ES cells were injected into C57Bl/6 embryos to generate chimeras. Chimeras were then crossed to C57Bl/6 mice in order to generate F1 offspring, which was confirmed by PCR.

8.3.1 g/3-21 mouse line expression/phenotype analysis

F1s positive for carrying the transposon insertion were crossed to a germ line *Cre* expressing line (*g/cre*) in order to utilise the presence of the *LoxP* sites within the targeting vector and generate an inversion encompassing the 5´ end of the L*mbr1* gene. So far no mice positive for the inversion have been born and in fact low pup numbers have been found consistently during crosses of 3-21 males with *Cre* female mice (data generated from 4 litters).

8.3.2 g/3-119 mouse line expression/phenotype analysis

ES cells positive for the SBLac3-119 insert were injected into C57Bl/6 embryos to generate chimeras. They were also treated with a *Cre recombinase in vitro* to generate a deletion within the cell line; this was confirmed by PCR and again used to generate chimeras. Of the chimeras generated only one, generated from the pre-deletion ES cells was found to be fertile. This chimera generated a low number of F1 offspring positive for the SBLac insert which were themselves runted and unable to breed (n=4/16 from 4 litters). It was therefore not possible to generate a deletion mouse line using this ES cell line.

8.4 Generation of *Lmbr1* knock-out line

8.4.1 Designing a targeting vector to 'Knock-out' Lmbr1

In order the complement the lmbr1_LHED work it was decided to design a targeting vector which could be used to delete the 5´ end of *Lmbr1*. A targeting vector was designed which when used in combination with Lmbr1_LHED targeted cells would allow the 'knocking-out' of the *Lmbr1* gene. This deletion would encompass the *Lmbr1* transcriptional start site as well as some upstream region thus allowing investigation to determine if *Lmbr1* transcriptional activity is required for *Zrs* activity.

The vector was designed to contain a 6.4 kb genomic arm homologous to a region approximately 20 kb upstream of *Lmbr1*. The vector was also designed so that a *LoxP* site as well as a hygromycin resistance gene could be inserted within the homology arm in an orientation so that when targeted alongside Lmbr1_LHED (which also contains *LoxP* sites) the 20 kb region upstream of *Lmbr1* could be deleted and the gene could be rendered transcriptionally inactive.

8.4.2 Vector Cloning

Primers were designed to amplify 300 bp regions from the 5´ and 3´ ends of the homology arm from BAC DNA encompassing the *Lmbr1* region (Figure 8.3A [i]). Primers were also designed to include restriction enzyme sites for cloning into pBluescript (*Not1/Cla1* or *Cla1/Sal1*). Ligations were then performed to clone the 5´ homology arm PCR fragment into the pBluescript vector and visualised by colony PCR. Following transformation, small-scale plasmid purification was performed on a select number of colonies and the resulting DNA was sequenced using M13 forward and reverse primers, which confirmed no mutations had been induced during the PCR. The vectors containing the 5´ end of the homology arm was then digested with *Not1* and *Cla1* and a ligation set up to ligate the 3´ end of the homology arm into the same vector. Following transformation restriction digest was used to identify vectors containing both ends of the homology arm (Figure 8.3B[i]).

In order to generate the full homology arm bacterial recombineering was employed. A double strand break was generated into the vector using the restriction

enzyme *Cla1* and the DNA was subsequently electroporated into electro-competent EL350 gap repair cells.

A mini-targeting vector containing a hygromycin resistance gene was designed to insert within the homology arm. PCR was used to generate 200 bp fragments homologous to sequence within the homology arm (Figure 8.3A [ii]). Primers were designed to allow the addition of restriction sites either side of the PCR fragments (*Not/Xho* or *Xho/Sal*). These fragments were then ligated into a pBluescript vector which already contained a Floxed Hygro Cassette (Figure 8.3B [ii]). Following completion of cloning the mini-targeting vector was linearised using *Not/Xho* and was electroporated alongside the targeting vector into EL350 recombinant cells (Figure 8.3C). Colony PCR using primers homologous to the hygromycin gene were used to screen the resulting colonies. Positive clones were then grown up and purified using small-scale plasmid purification before restriction digest was used to confirm correct cloning of the vector (Figure 8.3C).

8.4.3 ES cell Targeting

Following completion of cloning and sequence verification, the *Lmbr1* knock-out vector was linearised using *Sal1* and electroporated in E14 ES cells already targeted with the Lmbr1-pLHED vector. Cells were then selected for approximately 10 days using Hygromycin. Following selection, approximately 100 clones were picked and DNA was extracted from them. Long range PCR was employed to identify positively targeted clones using primers against the hygromycin cassette and out with the homology arms. One clone (out of 100 analysed) was found to have been correctly targeted (1%) (Figure 8.3D).

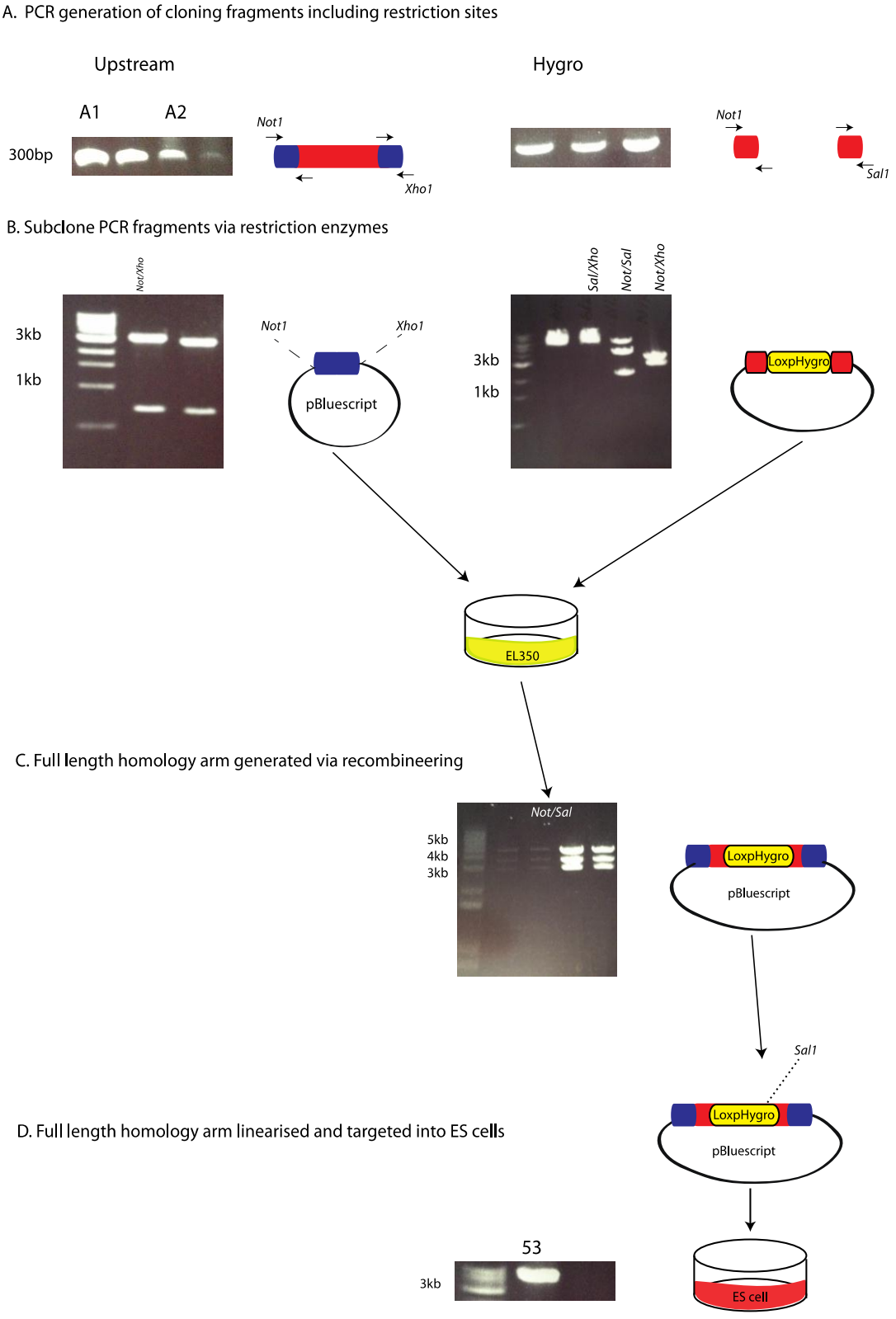


Figure 8.3: Cloning of an upstream Lmbr1 targeting vector

A. PCR fragments were generated for 300 bp homology arm ends (dark blue) (ii) and 200 bp internal fragments (red) (ii).
B. Upstream Lmbr1 Fragments were cloned into the pBluescript vector and confirmed by restriction digest with *Not1/Sal1* (i). Internal hygro fragments were cloned either side of a *loxP/hygro* cassette, confirmed by restriction digest (ii).
C. Both vectors were electroporated simultaneously into EL350 recombineering bacteria.
D. A pBluescript vector containing a complete homology arm including LoxP/hygro cassette was confirmed by restriction digest with *Not1/Sal1*.
E. The finished vector was linearised using *Sal1* and targeted into E14 ES cells, this was confirmed by PCR.

8.5.4 *Lmbr1* 'Knock-out' Mouse production

Cells positively targeted with both Lmbr1_LHED and Lmbr1 'Knock-out' vectors were used for blastocyst injection into C57BL6/J super ovulated females. Four chimeric mice were produced. These chimeras were then used to generate F1s by crossing them with C57BL6/J females. The resulting litter were genotyped by PCR using primers against *LacZ* (LHED insertion) and hygromycin (additional LoxP site). It became evident that during targeting the second insertion had actually inserted on the opposite chromosome. A cross was therefore set up instead to cross the LHED mice to *Cre* expressing females to generate mice positive for the insertion and *Cre recombinase*. These mice were then crossed to *LoxP* carrying mice in order to recombine the inserts onto the same chromosome and generate the deletion. Resulting pups from three litters (approximately n=34) were genotyped however it was not possible to identify any which had undergone the recombination of both *LoxP* sites onto the same chromosome.

8.6 Conclusions

It was shown to be possible to examine the chromosomal region surround *Shh* within ES cells. In wild-type ES cell*s Shh* and the *Zrs* were found to be held in a tighter conformation than expected despite the gene not being expressed. This phenomenon has also been previously observed within anterior limb bud cells, where an interaction is still seen between the *Zrs* an*d Shh* despit*e Shh* not being expressed (Amano et al. 2009). This is attributed to th*e Shh* locus

being held in one of three conformations, silent, poised or active (Amano et al. 2009). It is therefore suggested that within wild-type ES cells th*e Shh* locus is held in a poised position even thoug*h Shh* is not expressed. When the ES cells were treated with retinoic acid and Shh expression induced, the region was found to compact with very small interprobe distances found betwee*n Shh* and the *Zrs*. This is likely to reflect an active conformation.

Unlike what is reported with the literature (Amano et al. 2009) the FISH analysis does not suggest that *Shh* and the ZRS are experiencing 'looping'. If this was the case it would be expected that probes for regions within the middle of the locus would be found at larger distances from *Shh* than the ZRS. However, instead the whole region appears to be found at relatively small distances. This would suggest that instead the locus is actually just very compact and upon Shh expression it compacts further.

It was hoped that the region upstream of *Lmbr1* could be examined via deletion/inversion with a hope of disrupting the boundaries previously indicated by both already defined TADs and also tetraploid complementation mapping. Breakpoints within complex loci have been shown to result in overt developmental phenotypes; rearrangements of the *PAX6* locus have been shown to be associated with aniridia (Kleinjan and van Heyningen 1998; Kleinjan et al. 2001). Within the *Hox* cluster, any condition which interrupts this interval, by either an inversion or duplication intercalating some DNA sequences within the regulatory landscape, has been shown to result in a partial loss of expression. Thus suggesting, it is critical for the proper transcriptional control of genes as an integral regulatory block to be preserved (Montavon et al. 2011). Within both mice and humans rearrangements at the *Shh* locus have also been identified. One such human case was shown to result in a phenotype including a solitary maxillary central incisor (SMCI); a facial feature associated with HPE spectrum disorder as well as complete syndactyly of the hands and feet. This was expected to have occurred as a result of both gain- and loss-of-function *cis*-regulatory mechanisms. The HPE spectrum disorder in the child was attributed to the loss of *SHH* enhancers as a result of the inversion with a gain of

function “enhancer adoption” effect speculated as a cause of the preaxial synpolydactyly (Lettice et al. 2011). The removal of the boundary elements surrounding *Shh* would therefore be a way to examine how enhancers behave when they have the possibility of activating new genes and what happens when the genomic context of these enhancers is disturbed. Unfortunately as yet it has not been possible to produce mice with either deletions or inversions within this region. While the SBLac3-119 has resulted in only runted offspring as yet and it has not been possible to generate either deletions or inversions in this region, this could suggest a dominant phenotype which requires further analysis. However it is also possible that genetic issues are occurring within these mice unrelated to the various modifications generated within them. All three lines generated for examination of this region came from the L35 ES cell line (carrying the Lmbr1_LHED insertion vector), whereas the three lines used to generate the deletions between *Shh* and the *Zrs* came from the S32 line (Carrying the Shh2_LHED insertion vector). These cells underwent several rounds of targeting and selection and although no obvious chromosomal defects were identified it is possible subtle mutations have been acquired. Each of the lines is therefore currently going through several rounds of backcrossing before attempting to cross with Cre recombinase mice again.

Location of LHED reinsertions

Table 8.1: Reinsertion sites of the LHED transposon after SB100X mobilzation

Clones labelled with single numbers e.g. 120 represent those in which came from the original transfection within the S32 line, whereas the prefix '2' e.g. 2-25 is used to denote second transfection on S32 cells. Finally the prefix'3' e.g. 3-119 represents clones harvested following the single transfection within the L35 cells.

Clone Numbers	Chromosome	Start site
3-63	1	10132129
2-28	1	13035536
28	1	64817966
2-56	1	115299424
30	1	151412958
152	2	9776384
2-19	2	52678558
2-159	2	133268860
2-132	2	166876113
3-10	3	3794132
3-26	3	4866073
3-55	3	4866073
124	3	19862883
3-31	3	92016171
3-123	4	14797281
2-30	4	61875857
2-149	4	85868154
2-190	4	85868195
2-17	4	116259347
3-56	4	135914822
3-91	4	139320332
3-113	5	9168148
175	5	9233859
183	5	9235918
191	5	9235918
2-84	5	11836765
35	5	12226155
2-79	5	12469518
3-104	5	12781420
2-33	5	13565636
119	5	14780967
71	5	14952294
3-124	5	15136713
169	5	15196170
3-69	5	15201982
3-102	5	16581331
194	5	16826233

5	5	19554861
181	5	19554861
6	5	20477133
2	5	21271005
2-107	5	21740196
2-101	5	23262039
197	5	23882414
2-83	5	24419635
31	5	26381308
3-19	5	26425225
2-76	5	26425323
11	5	26427887
75	5	26617860
114	5	27059363
3-134	5	27193105
97	5	27351309
2-130	5	27541737
2-139	5	28141761
3-2	5	28483004
138	5	28500403
66	5	28505362
3-24	5	28684811
2-110	5	28687701
3-99	5	28777670
3-41	5	28777672
7	5	28787130
3-70	5	28790598
2-51	5	28796870
98	5	28797890
2-4	5	28889281
120	5	28889622
2-39	5	28889761
115	5	28978743
109	5	28993824
110	5	28993824
3-93	5	28996311
3-96	5	29002839
2-88	5	29011233
148	5	29066757
46	5	29116092
2-77	5	29266136
131	5	29278627
2-67	5	29319991
3-76	5	29403698
2-73	5	29405968

2-148	5	29443264
83	5	29444355
92	5	29444355
13	5	29445046
2-191	5	29468681
2-112	5	29471537
2-161	5	29471538
2-179	5	29475308
2-120	5	29489431
3-5	5	29498108
2-134	5	29515622
2-25	5	29535419
130	5	29583183
2-200	5	29589741
2-157	5	29591086
104	5	29622149
2-153	5	29622343
3-35	5	29626632
3-137	5	29626757
102	5	29630052
118	5	29634908
135	5	29635417
68	5	29639709
198	5	29647904
2-95	5	29648626
2-21	5	29648996
3-109	5	29649217
3-139	5	29653746
3-92	5	29653752
3-97	5	29654566
3-83	5	29660909
2-75	5	29722396
3 -21	5	29729576
3-60	5	29794631
2-32	5	29797324
3-119	5	29809901
156	5	29870513
26	5	30084989
3-133	5	30252588
100	5	30412905
3-47	5	30417140
3-20	5	30564372
44	5	31017172
2-187	5	31032905
99	5	31754503

17	5	32175629
132	5	32491259
3-38	5	33985251
43	5	35312641
3-65	5	36904719
128	5	38227932
189	5	41093593
112	5	42993672
2-121	5	43026925
111	5	44076759
2-150	5	45800705
2-127	5	47579156
3-40	5	48010706
3-53	5	48015632
133	5	48229228
3-15	5	48446271
78	5	49299399
157	5	49972984
3 42	5	50348005
3-7	5	50348122
2-7	5	50644953
9	5	51550601
2-89	5	51797701
3-95	5	52538277
3-36	5	52538747
2-111	5	53976070
3-18	5	55701651
2-146	5	57116619
36	5	57140720
150	5	58426076
2-100	5	63964545
199	5	64758091
2-64	5	65066713
2-98	5	65398770
58	5	66700503
2-155	5	69418387
34	5	69934611
3-85	5	71399112
3-121	5	73217101
3-48	5	76672996
3-51	5	76672996
108	5	78241139
2-16	5	91598106
2-59	5	91868398
171	5	93230434

3-80	5	96192364
2-3	5	98836299
168	5	114367950
3-94	5	120161311
136	5	128326269
3-115	5	135019134
3-30	5	29798342
188	6	9790272
2-105	6	33574026
96	6	113059174
89	7	7644373
3-22	7	13376563
3	7	19820575
2-128	7	20472003
2-24	7	62375427
2-129	7	62375431
3-78	7	83068342
2-90	7	120715376
3-111	8	72271096
122	8	73217921
2-22	8	131461255
3-117	9	752036
2-69	9	7224628
3-57	9	9418548
3-28	9	105239133
86	10	5365227
177	10	7113602
3-114	10	57165077
2-117	10	77538788
2-182	10	93653604
3-77	10	119604549
3-39	11	4334449
16	11	62594873
3-32	11	77999344
3-138	11	77999344
2-14	11	81915010
2-163	11	90377302
3-88	11	93579103
3-59	11	93579185
2-74	11	93924030
2-118	11	97868120
29	12	17088375
2-172	12	20891160
3-13	12	37708714
64	12	77343914

3-87	13	29361265
3-106	13	32005058
170	13	38285963
3 54	13	59837459
2-1	13	529797324
2-174	14	7452161
159	14	9609204
3-25	14	44752082
2-70	14	74864534
2-192	14	79903521
33	14	120203248
3-73	15	34130752
2-5	15	39536956
3-6	15	49690384
22	15	80653664
101	16	31121059
3-27	17	3181968
2-109	17	43957168
3-50	17	74968608
3-127	18	15181506
3-44	18	68612645
134	18	88798068
178	19	32105143
2-145	19	49106541
76	19	57183357
2-18	X	65899132
2-50	X	101579793

9. Acknowledgments

There are a number of people I have to thank for helping me get to the end of this study and completing this thesis.

Firstly Bob, for allowing me to work in his lab and for giving me what turned out to be an excellent project. I also have to give a great deal of thanks to Laura who has I think been just as invested in this project as I have and whose expertise and only minor exasperation I know I would not have managed without. Paul also deserves a lot of credit for the masses of mouse work he did for me, as well as everyone in the transgenic unit especially Gary for looking after my many mouse lines. I also want to thank Fiona and Joan for all their ES cell know-how which was so important in getting me through all my early failed transfections and contamination woes.

I also want to extend my thanks to everyone else on C3 especially those within both the Hill lab and also office mates who have provided lots of coffee, biscuits and moral support particularly during the writing process.

I also want to thank all my friends and family with special note to potentially the proudest parents in the world, this achievement is just as much theirs as it is mine.

Finally; Paula, thanks for loving dead mice almost as much as I do.

References

Ahlgren, S. C. and Bronner-Fraser, M. 1999. Inhibition of sonic hedgehog signaling in vivo results in craniofacial neural crest cell death. *Curr Biol* 9 (22):1304-14.

Ahn, S. and Joyner, A. L. 2004. Dynamic changes in the response of cells to positive hedgehog signaling during mouse limb patterning. *Cell* 118 (4):505-16.

Akiyama, H., Chaboissier, M. C., Martin, J. F., Schedl, A., and de Crombrugghe, B. 2002. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev* 16 (21):2813-28.

Allard, P. and Tabin, C. J. 2009. Achieving bilateral symmetry during vertebrate limb development. *Semin Cell Dev Biol* 20 (4):479-84.

Allen, B. L., Song, J. Y., Izzi, L., Althaus, I. W., Kang, J. S., Charron, F., Krauss, R. S., and McMahon, A. P. 2011. Overlapping roles and collective requirement for the coreceptors GAS1, CDO, and BOC in SHH pathway function. *Dev Cell* 20 (6):775-87.

Allen, B. L., Tenzen, T., and McMahon, A. P. 2007. The Hedgehog-binding proteins Gas1 and Cdo cooperate to positively regulate Shh signaling during mouse development. *Genes Dev* 21 (10):1244-57.

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. 2009. Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev Cell* 16 (1):47-57.

Anderson, E., Peluso, S., Lettice, L. A., and Hill, R. E. 2012. Human limb abnormalities caused by disruption of hedgehog signaling. *Trends Genet* 28 (8):364-73.

Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., Trono, D., Spitz, F., and Duboule, D. 2013. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* 340 (6137):1234167.

Araki, K., Araki, M., Miyazaki, J., and Vassalli, P. 1995. Site-specific recombination of a transgene in fertilized eggs by transient expression of Cre recombinase. *Proc Natl Acad Sci U S A* 92 (1):160-4.

Bagheri-Fam, S., Barrionuevo, F., Dohrmann, U., Gunther, T., Schule, R., Kemler, R., Mallo, M., Kanzler, B., and Scherer, G. 2006. Long-range upstream and downstream enhancers control distinct subsets of the complex spatiotemporal Sox9 expression pattern. *Dev Biol* 291 (2):382-97.

Bagheri-Fam, S., Ferraz, C., Demaille, J., Scherer, G., and Pfeifer, D. 2001. Comparative genomics of the SOX9 region in human and Fugu rubripes: conservation of short regulatory sequence elements within large intergenic regions. *Genomics* 78 (1-2):73-82.

Bakker, E. R., Das, A. M., Helvensteijn, W., Franken, P. F., Swagemakers, S., van der Valk, M. A., ten Hagen, T. L., Kuipers, E. J., van Veelen, W., and Smits, R. 2013. Wnt5a promotes human colon cancer cell migration and invasion but does not augment intestinal tumorigenesis in Apc1638N mice. *Carcinogenesis* 34 (11):2629-38.

Balczerski, B., Zakaria, S., Tucker, A. S., Borycki, A. G., Koyama, E., Pacifici, M., and Francis-West, P. 2012. Distinct spatiotemporal roles of hedgehog signalling during chick and mouse cranial base and axial skeleton development. *Dev Biol* 371 (2):203-14.

Beermann, F., Kaloulis, K., Hofmann, D., Murisier, F., Bucher, P., and Trumpp, A. 2006. Identification of evolutionarily conserved regulatory elements in the mouse Fgf8 locus. *Genesis* 44 (1):1-6.

Belloni, E., Muenke, M., Roessler, E., Traverso, G., Siegel-Bartelt, J., Frumkin, A., Mitchell, H. F., Donis-Keller, H., Helms, C., Hing, A. V., Heng, H. H., Koop, B., Martindale,

D., Rommens, J. M., Tsui, L. C., and Scherer, S. W. 1996. Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly. *Nat Genet* 14 (3):353-6.

Benko, S., Fantes, J. A., Amiel, J., Kleinjan, D. J., Thomas, S., Ramsay, J., Jamshidi, N., Essafi, A., Heaney, S., Gordon, C. T., McBride, D., Golzio, C., Fisher, M., Perry, P., Abadie, V., Ayuso, C., Holder-Espinasse, M., Kilpatrick, N., Lees, M. M., Picard, A., Temple, I. K., Thomas, P., Vazquez, M. P., Vekemans, M., Roest Crollius, H., Hastie, N. D., Munnich, A., Etchevers, H. C., Pelet, A., Farlie, P. G., Fitzpatrick, D. R., and Lyonnet, S. 2009. Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nat Genet* 41 (3):359-64.

Bhatia, N., Thiyagarajan, S., Elcheva, I., Saleem, M., Dlugosz, A., Mukhtar, H., and Spiegelman, V. S. 2006. Gli2 is targeted for ubiquitination and degradation by beta-TrCP ubiquitin ligase. *J Biol Chem* 281 (28):19320-6.

Bhatia, S., Monahan, J., Ravi, V., Gautier, P., Murdoch, E., Brenner, S., van Heyningen, V., Venkatesh, B., and Kleinjan, D. A. 2014. A survey of ancient conserved non-coding elements in the PAX6 locus reveals a landscape of interdigitated cis-regulatory archipelagos. *Dev Biol* 387 (2):214-28.

Bi, W., Deng, J. M., Zhang, Z., Behringer, R. R., and de Crombrugghe, B. 1999. Sox9 is required for cartilage formation. *Nat Genet* 22 (1):85-9.

Binns, W., Thacker, E. J., James, L. F., and Huffman, W. T. 1959. A congenital cyclopiantype malformation in lambs. *J Am Vet Med Assoc* 134 (4):180-3.

Birney, E.Stamatoyannopoulos, J. A.Dutta, A.Guigo, R.Gingeras, T. R.Margulies, E. H.Weng, Z.Snyder, M.Dermitzakis, E. T.Thurman, R. E.Kuehn, M. S.Taylor, C. M.Neph, S.Koch, C. M.Asthana, S.Malhotra, A.Adzhubei, I.Greenbaum, J. A.Andrews, R. M.Flicek, P.Boyle, P. J.Cao, H.Carter, N. P.Clelland, G. K.Davis, S.Day, N.Dhami, P.Dillon, S. C.Dorschner, M. O.Fiegler, H.Giresi, P. G.Goldy, J.Hawrylycz, M.Haydock, A.Humbert, R.James, K. D.Johnson, B. E.Johnson, E. M.Frum, T. T.Rosenzweig, E. R.Karnani, N.Lee, K.Lefebvre, G. C.Navas, P. A.Neri, F.Parker, S. C.Sabo, P. J.Sandstrom, R.Shafer, A.Vetrie, D.Weaver, M.Wilcox, S.Yu, M.Collins, F. S.Dekker, J.Lieb, J. D.Tullius, T. D.Crawford, G. E.Sunyaev, S.Noble, W. S.Dunham, I.Denoeud, F.Reymond, A.Kapranov, P.Rozowsky, J.Zheng, D.Castelo, R.Frankish, A.Harrow, J.Ghosh, S.Sandelin, A.Hofacker, I. L.Baertsch, R.Keefe, D.Dike, S.Cheng, J.Hirsch, H. A.Sekinger, E. A.Lagarde, J.Abril, J. F.Shahab, A.Flamm, C.Fried, C.Hackermuller, J.Hertel, J.Lindemeyer, M.Missal, K.Tanzer, A.Washietl, S.Korbel, J.Emanuelsson, O.Pedersen, J. S.Holroyd, N.Taylor, R.Swarbreck, D.Matthews, N.Dickson, M. C.Thomas, D. J.Weirauch, M. T.Gilbert, J.Drenkow, J.Bell, I.Zhao, X.Srinivasan, K. G.Sung, W. K.Ooi, H. S.Chiu, K. P.Foissac, S.Alioto, T.Brent, M.Pachter, L.Tress, M. L.Valencia, A.Choo, S. W.Choo, C. Y.Ucla, C.Manzano, C.Wyss, C.Cheung, E.Clark, T. G.Brown, J. B.Ganesh, M.Patel, S.Tammana, H.Chrast, J.Henrichsen, C. N.Kai, C.Kawai, J.Nagalakshmi, U.Wu, J.Lian, Z.Lian, J.Newburger, P.Zhang, X.Bickel, P.Mattick, J. S.Carninci, P.Hayashizaki, Y.Weissman, S.Hubbard, T.Myers, R. M.Rogers, J.Stadler, P. F.Lowe, T. M.Wei, C. L.Ruan, Y.Struhl, K.Gerstein, M.Antonarakis, S. E.Fu, Y.Green, E. D.Karaoz, U.Siepel, A.Taylor, J.Liefer, L. A.Wetterstrand, K. A.Good, P. J.Feingold, E. A.Guyer, M. S.Cooper, G. M.Asimenos, G.Dewey, C. N.Hou, M.Nikolaev, S.Montoya-Burgos, J. I.Loytynoja, A.Whelan, S.Pardi, F.Massingham, T.Huang, H.Zhang, N. R.Holmes, I.Mullikin, J. C.Ureta-Vidal, A.Paten, B.Seringhaus, M.Church, D.Rosenbloom, K.Kent, W. J.Stone, E. A.Batzoglou, S.Goldman, N.Hardison, R. C.Haussler, D.Miller, W.Sidow, A.Trinklein, N. D.Zhang, Z. D.Barrera, L.Stuart, R.King, D. C.Ameur, A.Enroth, S.Bieda, M. C.Kim, J.Bhinge, A. A.Jiang, N.Liu, J.Yao, F.Vega, V. B.Lee, C. W.Ng, P.Yang, A.Moqtaderi, Z.Zhu, Z.Xu, X.Squazzo, S.Oberley, M. J.Inman, D.Singer, M. A.Richmond, T. A.Munn, K. J.Rada-Iglesias, A.Wallerman,

O.Komorowski, J.Fowler, J. C.Couttet, P.Bruce, A. W.Dovey, O. M.Ellis, P. D.Langford, C. F.Nix, D. A.Euskirchen, G.Hartman, S.Urban, A. E.Kraus, P.Van Calcar, S.Heintzman, N.Kim, T. H.Wang, K.Qu, C.Hon, G.Luna, R.Glass, C. K.Rosenfeld, M. G.Aldred, S. F.Cooper, S. J.Halees, A.Lin, J. M.Shulha, H. P.Xu, M.Haidar, J. N.Yu, Y.Iyer, V. R.Green, R. D.Wadelius, C.Farnham, P. J.Ren, B.Harte, R. A.Hinrichs, A. S.Trumbower, H.Clawson, H.Hillman-Jackson, J.Zweig, A. S.Smith, K.Thakkapallayil, A.Barber, G.Kuhn, R. M.Karolchik, D.Armengol, L.Bird, C. P.de Bakker, P. I.Kern, A. D.Lopez-Bigas, N.Martin, J. D.Stranger, B. E.Woodroffe, A.Davydov, E.Dimas, A.Eyras, E.Hallgrimsdottir, I. B.Huppert, J.Zody, M. C.Abecasis, G. R.Estivill, X.Bouffard, G. G.Guan, X.Hansen, N. F.Idol, J. R.Maduro, V. V.Maskeri, B.McDowell, J. C.Park, M.Thomas, P. J.Young, A. C.Blakesley, R. W.Muzny, D. M.Sodergren, E.Wheeler, D. A.Worley, K. C.Jiang, H.Weinstock, G. M.Gibbs, R. A.Graves, T.Fulton, R.Mardis, E. R.Wilson, R. K.Clamp, M.Cuff, J.Gnerre, S.Jaffe, D. B.Chang, J. L.Lindblad-Toh, K.Lander, E. S.Koriabine, M.Nefedov, M.Osoegawa, K.Yoshinaga, Y.Zhu, B. and de Jong, P. J. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447 (7146):799-816.

Bitgood, M. J. and McMahon, A. P. 1995. Hedgehog and Bmp genes are coexpressed at many diverse sites of cell-cell interaction in the mouse embryo. *Dev Biol* 172 (1):126-38.

Bitgood, M. J., Shen, L., and McMahon, A. P. 1996. Sertoli cell signaling by Desert hedgehog regulates the male germline. *Curr Biol* 6 (3):298-304.

Botchkarev, V. A., Botchkareva, N. V., Sharov, A. A., Funa, K., Huber, O., and Gilchrest, B. A. 2002. Modulation of BMP signaling by noggin is required for induction of the secondary (nontylotrich) hair follicles. *J Invest Dermatol* 118 (1):3-10.

Brito, J. M., Teillet, M. A., and Le Douarin, N. M. 2006. An early role for sonic hedgehog from foregut endoderm in jaw development: ensuring neural crest cell survival. *Proc Natl Acad Sci U S A* 103 (31):11607-12.

———. 2008. Induction of mirror-image supernumerary jaws in chicken mandibular mesenchyme by Sonic Hedgehog-producing cells. *Development* 135 (13):2311-9.

Britto, J., Tannahill, D., and Keynes, R. 2002. A critical role for sonic hedgehog signaling in the early expansion of the developing brain. *Nat Neurosci* 5 (2):103-10.

Bulger, M., Schubeler, D., Bender, M. A., Hamilton, J., Farrell, C. M., Hardison, R. C., and Groudine, M. 2003. A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse beta-globin locus. *Mol Cell Biol* 23 (15):5234-44.

Burke, D. T., Carle, G. F., and Olson, M. V. 1987. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236 (4803):806-12.

Bushey, A. M., Dorman, E. R., and Corces, V. G. 2008. Chromatin insulators: regulatory mechanisms and epigenetic inheritance. *Mol Cell* 32 (1):1-9.

Butler, J. E. and Kadonaga, J. T. 2001. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev* 15 (19):2515-9.

Cajiao, I., Zhang, A., Yoo, E. J., Cooke, N. E., and Liebhaber, S. A. 2004. Bystander gene activation by a locus control region. *EMBO J* 23 (19):3854-63.

Calhoun, V. C., Stathopoulos, A., and Levine, M. 2002. Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. *Proc Natl Acad Sci U S A* 99 (14):9243-7.

Celniker, S. E. and Drewell, R. A. 2007. Chromatin looping mediates boundary element promoter interactions. *Bioessays* 29 (1):7-10.

Chaboissier, M. C., Kobayashi, A., Vidal, V. I., Lutzkendorf, S., van de Kant, H. J., Wegner, M., de Rooij, D. G., Behringer, R. R., and Schedl, A. 2004. Functional analysis of

Sox8 and Sox9 during sex determination in the mouse. *Development* 131 (9):1891-901.
Chang, D. T., Lopez, A., von Kessler, D. P., Chiang, C., Simandl, B. K., Zhao, R., Seldin, M. F., Fallon, J. F., and Beachy, P. A. 1994. Products, genetic linkage and limb patterning activity of a murine hedgehog gene. *Development* 120 (11):3339-53.
Chen, M. H., Wilson, C. W., Li, Y. J., Law, K. K., Lu, C. S., Gacayan, R., Zhang, X., Hui, C. C., and Chuang, P. T. 2009. Cilium-independent regulation of Gli protein function by Sufu in Hedgehog signaling is evolutionarily conserved. *Genes Dev* 23 (16):1910-28.
Chi, C. L., Martinez, S., Wurst, W., and Martin, G. R. 2003. The isthmic organizer signal FGF8 is required for cell survival in the prospective midbrain and cerebellum. *Development* 130 (12):2633-44.
Chiang, C., Litingtung, Y., Lee, E., Young, K. E., Corden, J. L., Westphal, H., and Beachy, P. A. 1996. Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* 383 (6599):407-13.
Chong, H. J., Young, N. M., Hu, D., Jeong, J., McMahon, A. P., Hallgrimsson, B., and Marcucio, R. S. 2012. Signaling by SHH rescues facial defects following blockade in the brain. *Dev Dyn* 241 (2):247-56.
Chuang, P. T. and McMahon, A. P. 1999. Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. *Nature* 397 (6720):617-21.
Chung, J. H., Whiteley, M., and Felsenfeld, G. 1993. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell* 74 (3):505-14.
Chuong, C. M., Patel, N., Lin, J., Jung, H. S., and Widelitz, R. B. 2000. Sonic hedgehog signaling pathway in vertebrate epithelial appendage morphogenesis: perspectives in development and evolution. *Cell Mol Life Sci* 57 (12):1672-81.
Clark, R. M., Marker, P. C., and Kingsley, D. M. 2000. A novel candidate gene for mouse and human preaxial polydactyly with altered expression in limbs of Hemimelic extra-toes mutant mice. *Genomics* 67 (1):19-27.
Clark, R. M., Marker, P. C., Roessler, E., Dutra, A., Schimenti, J. C., Muenke, M., and Kingsley, D. M. 2001. Reciprocal mouse and human limb phenotypes caused by gain- and loss-of-function mutations affecting Lmbr1. *Genetics* 159 (2):715-26.
Cobourne, M. T., Miletich, I., and Sharpe, P. T. 2004. Restriction of sonic hedgehog signalling during early tooth development. *Development* 131 (12):2875-85.
Consortium, T. E. P. 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* 9 (4):e1001046.
Cooper, M. K., Porter, J. A., Young, K. E., and Beachy, P. A. 1998. Teratogen-mediated inhibition of target tissue response to Shh signaling. *Science* 280 (5369):1603-7.
Corbin, J. G., Nery, S., and Fishell, G. 2001. Telencephalic cells take a tangent: non-radial migration in the mammalian forebrain. *Nat Neurosci* 4 Suppl:1177-82.
Corbin, J. G., Rutlin, M., Gaiano, N., and Fishell, G. 2003. Combinatorial function of the homeodomain proteins Nkx2.1 and Gsh2 in ventral telencephalic patterning. *Development* 130 (20):4895-906.
Corbit, K. C., Aanstad, P., Singla, V., Norman, A. R., Stainier, D. Y., and Reiter, J. F. 2005. Vertebrate Smoothened functions at the primary cilium. *Nature* 437 (7061):1018-21.
Cordero, D., Marcucio, R., Hu, D., Gaffield, W., Tapadia, M., and Helms, J. A. 2004. Temporal perturbations in sonic hedgehog signaling elicit the spectrum of holoprosencephaly phenotypes. *J Clin Invest* 114 (4):485-94.
Crackower, M. A., Scherer, S. W., Rommens, J. M., Hui, C. C., Poorkaj, P., Soder, S., Cobben, J. M., Hudgins, L., Evans, J. P., and Tsui, L. C. 1996. Characterization of the split hand/split foot malformation locus SHFM1 at 7q21.3-q22.1 and analysis of a candidate gene for its expression during limb development. *Hum Mol Genet* 5 (5):571-9.

Davis, C. A., Noble-Topham, S. E., Rossant, J., and Joyner, A. L. 1988. Expression of the homeo box-containing gene En-2 delineates a specific region of the developing mouse brain. *Genes Dev* 2 (3):361-71.

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., and Ren, B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485 (7398):376-80.

Duverger, O. and Morasso, M. I. 2009. Epidermal patterning and induction of different hair types during mouse embryonic development. *Birth Defects Res C Embryo Today* 87 (3):263-72.

Eaton, S. 2008. Multiple roles for lipids in the Hedgehog signalling pathway. *Nat Rev Mol Cell Biol* 9 (6):437-45.

Echelard, Y., Epstein, D. J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J. A., and McMahon, A. P. 1993. Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell* 75 (7):1417-30.

ENCODE. 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* 9 (4):e1001046.

Epstein, D. J., McMahon, A. P., and Joyner, A. L. 1999. Regionalization of Sonic hedgehog transcription along the anteroposterior axis of the mouse central nervous system is regulated by Hnf3-dependent and -independent mechanisms. *Development* 126 (2):281-92.

Ericson, J., Briscoe, J., Rashbass, P., van Heyningen, V., and Jessell, T. M. 1997. Graded sonic hedgehog signaling and the specification of cell fate in the ventral neural tube. *Cold Spring Harb Symp Quant Biol* 62:451-66.

Ericson, J., Morton, S., Kawakami, A., Roelink, H., and Jessell, T. M. 1996. Two critical periods of Sonic Hedgehog signaling required for the specification of motor neuron identity. *Cell* 87 (4):661-73.

Fan, C. M. and Tessier-Lavigne, M. 1994. Patterning of mammalian somites by surface ectoderm and notochord: evidence for sclerotome induction by a hedgehog homolog. *Cell* 79 (7):1175-86.

Farrell, C. M., West, A. G., and Felsenfeld, G. 2002. Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Mol Cell Biol* 22 (11):3820-31.

Fedoriw, A. M., Stein, P., Svoboda, P., Schultz, R. M., and Bartolomei, M. S. 2004. Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. *Science* 303 (5655):238-40.

Fisher, S., Grice, E. A., Vinton, R. M., Bessling, S. L., and McCallion, A. S. 2006. Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science* 312 (5771):276-9.

Foster, J. W., Dominguez-Steglich, M. A., Guioli, S., Kwok, C., Weller, P. A., Stevanovic, M., Weissenbach, J., Mansour, S., Young, I. D., Goodfellow, P. N., and et al. 1994. Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature* 372 (6506):525-30.

Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F., and Stern, D. L. 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* 466 (7305):490-3.

Fuccillo, M., Rallu, M., McMahon, A. P., and Fishell, G. 2004. Temporal requirement for hedgehog signaling in ventral telencephalic patterning. *Development* 131 (20):5031-40.

Galli, A., Robay, D., Osterwalder, M., Bao, X., Benazet, J. D., Tariq, M., Paro, R., Mackem, S., and Zeller, R. 2010. Distinct roles of Hand2 in initiating polarity and posterior Shh expression during the onset of mouse limb bud development. *PLoS Genet* 6 (4):e1000901.

Gammill, L. S. and Bronner-Fraser, M. 2003. Neural crest specification: migrating into genomics. *Nat Rev Neurosci* 4 (10):795-805.

Garel, S., Huffman, K. J., and Rubenstein, J. L. 2003. Molecular regionalization of the neocortex is disrupted in Fgf8 hypomorphic mutants. *Development* 130 (9):1903-14.
Gazzerro, E., Pereira, R. C., Jorgetti, V., Olson, S., Economides, A. N., and Canalis, E. 2005. Skeletal overexpression of gremlin impairs bone formation and causes osteopenia. *Endocrinology* 146 (2):655-65.
George, S. H., Gertsenstein, M., Vintersten, K., Korets-Smith, E., Murphy, J., Stevens, M. E., Haigh, J. J., and Nagy, A. 2007. Developmental and adult phenotyping directly from mutant embryonic stem cells. *Proc Natl Acad Sci U S A* 104 (11):4455-60.
Gerasimova, T. I. and Corces, V. G. 2001. Chromatin insulators and boundaries: effects on transcription and nuclear organization. *Annu Rev Genet* 35:193-208.
Geyer, P. K. 1997. The role of insulator elements in defining domains of gene expression. *Curr Opin Genet Dev* 7 (2):242-8.
Glaser, T., Jepeal, L., Edwards, J. G., Young, S. R., Favor, J., and Maas, R. L. 1994. PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. *Nat Genet* 7 (4):463-71.
Gonzalez, F., Duboule, D., and Spitz, F. 2007. Transgenic analysis of Hoxd gene regulation during digit development. *Dev Biol* 306 (2):847-59.
Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R., and Fraser, P. 2000. Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. *Mol Cell* 5 (2):377-86.
Gritli-Linde, A., Lewis, P., McMahon, A. P., and Linde, A. 2001. The whereabouts of a morphogen: direct evidence for short- and graded long-range activity of hedgehog signaling peptides. *Dev Biol* 236 (2):364-86.
Grove, E. A., Tole, S., Limon, J., Yip, L., and Ragsdale, C. W. 1998. The hem of the embryonic cerebral cortex is defined by the expression of multiple Wnt genes and is compromised in Gli3-deficient mice. *Development* 125 (12):2315-25.
Gunhaga, L., Jessell, T. M., and Edlund, T. 2000. Sonic hedgehog signaling at gastrula stages specifies ventral telencephalic cells in the chick embryo. *Development* 127 (15):3283-93.
Hadzhiev, Y., Lang, M., Ertzer, R., Meyer, A., Strahle, U., and Muller, F. 2007. Functional diversification of sonic hedgehog paralog enhancers identified by phylogenomic reconstruction. *Genome Biol* 8 (6):R106.
Hall, B., Limaye, A., and Kulkarni, A. B. 2009. Overview: generation of gene knockout mice. *Curr Protoc Cell Biol* Chapter 19:Unit 19 12 19 12 1-17.
Hall, J. M., Hooper, J. E., and Finger, T. E. 1999. Expression of sonic hedgehog, patched, and Gli1 in developing taste papillae of the mouse. *J Comp Neurol* 406 (2):143-55.
Hall, T. M., Porter, J. A., Beachy, P. A., and Leahy, D. J. 1995. A potential catalytic site revealed by the 1.7-A crystal structure of the amino-terminal signalling domain of Sonic hedgehog. *Nature* 378 (6553):212-6.
Han, Y. G., Kwok, B. H., and Kernan, M. J. 2003. Intraflagellar transport is required in Drosophila to differentiate sensory cilia but not sperm. *Curr Biol* 13 (19):1679-86.
Harfe, B. D., Scherz, P. J., Nissim, S., Tian, H., McMahon, A. P., and Tabin, C. J. 2004. Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. *Cell* 118 (4):517-28.
Heath, H., Ribeiro de Almeida, C., Sleutels, F., Dingjan, G., van de Nobelen, S., Jonkers, I., Ling, K. W., Gribnau, J., Renkawitz, R., Grosveld, F., Hendriks, R. W., and Galjart, N. 2008. CTCF regulates cell cycle progression of alphabeta T cells in the thymus. *EMBO J* 27 (21):2839-50.
Helms, J. A. and Schneider, R. A. 2003. Cranial skeletal biology. *Nature* 423 (6937):326-31.
Herault, Y. and Duboule, D. 1996. [Genetic control of limb development]. *Ann Genet* 39 (4):222-32.
Heus, H. C., Hing, A., van Baren, M. J., Joosse, M., Breedveld, G. J., Wang, J. C., Burgess, A., Donnis-Keller, H., Berglund, C., Zguricas, J., Scherer, S. W., Rommens, J. M.,

Oostra, B. A., and Heutink, P. 1999. A physical and transcriptional map of the preaxial polydactyly locus on chromosome 7q36. *Genomics* 57 (3):342-51.
Hill, R. E. 2007. How to make a zone of polarizing activity: insights into limb development via the abnormality preaxial polydactyly. *Dev Growth Differ* 49 (6):439-48.
Hirate, Y. and Okamoto, H. 2006. Canopy1, a novel regulator of FGF signaling around the midbrain-hindbrain boundary in zebrafish. *Curr Biol* 16 (4):421-7.
Hobert, O. 2010. Gene regulation: enhancers stepping out of the shadow. *Curr Biol* 20 (17):R697-9.
Holm, I., Monclair, T., Lundar, T., Stadheim, B., Prescott, T. E., and Eiklid, K. L. 2013. A 5.8 kb deletion removing the entire MNX1 gene in a Norwegian family with Currarino syndrome. *Gene* 518 (2):457-60.
Holmes, G., Crooijmans, R., Groenen, M., and Niswander, L. 2003. ALC (adjacent to LMX1 in chick) is a novel dorsal limb mesenchyme marker. *Gene Expr Patterns* 3 (6):735-41.
Hou, C., Li, L., Qin, Z. S., and Corces, V. G. 2012. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol Cell* 48 (3):471-84.
Houston, C. S., Opitz, J. M., Spranger, J. W., Macpherson, R. I., Reed, M. H., Gilbert, E. F., Herrmann, J., and Schinzel, A. 1983. The campomelic syndrome: review, report of 17 cases, and follow-up on the currently 17-year-old boy first reported by Maroteaux et al in 1971. *Am J Med Genet* 15 (1):3-28.
Huangfu, D. and Anderson, K. V. 2005. Cilia and Hedgehog responsiveness in the mouse. *Proc Natl Acad Sci U S A* 102 (32):11325-30.
———. 2006. Signaling from Smo to Ci/Gli: conservation and divergence of Hedgehog pathways from Drosophila to vertebrates. *Development* 133 (1):3-14.
Incardona, J. P., Gaffield, W., Kapur, R. P., and Roelink, H. 1998. The teratogenic Veratrum alkaloid cyclopamine inhibits sonic hedgehog signal transduction. *Development* 125 (18):3553-62.
Ingham, P. W. 1993. Localized hedgehog activity controls spatial limits of wingless transcription in the Drosophila embryo. *Nature* 366 (6455):560-2.
Iseki, S., Araga, A., Ohuchi, H., Nohno, T., Yoshioka, H., Hayashi, F., and Noji, S. 1996. Sonic hedgehog is expressed in epithelial cells during development of whisker, hair, and tooth. *Biochem Biophys Res Commun* 218 (3):688-93.
Jeong, J., Mao, J., Tenzen, T., Kottmann, A. H., and McMahon, A. P. 2004. Hedgehog signaling in the neural crest cells regulates the patterning and growth of facial primordia. *Genes Dev* 18 (8):937-51.
Jeong, Y., Dolson, D. K., Waclaw, R. R., Matise, M. P., Sussel, L., Campbell, K., Kaestner, K. H., and Epstein, D. J. Spatial and temporal requirements for sonic hedgehog in the regulation of thalamic interneuron identity. *Development* 138 (3):531-41.
———. 2011. Spatial and temporal requirements for sonic hedgehog in the regulation of thalamic interneuron identity. *Development* 138 (3):531-41.
Jeong, Y., El-Jaick, K., Roessler, E., Muenke, M., and Epstein, D. J. 2006. A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. *Development* 133 (4):761-72.
Jeong, Y., Leskow, F. C., El-Jaick, K., Roessler, E., Muenke, M., Yocum, A., Dubourg, C., Li, X., Geng, X., Oliver, G., and Epstein, D. J. 2008. Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nat Genet* 40 (11):1348-53.
Jessell, T. M. 2000. Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nat Rev Genet* 1 (1):20-9.
Jia, J., Amanai, K., Wang, G., Tang, J., Wang, B., and Jiang, J. 2002. Shaggy/GSK3 antagonizes Hedgehog signalling by regulating Cubitus interruptus. *Nature* 416 (6880):548-52.

Jia, J., Zhang, L., Zhang, Q., Tong, C., Wang, B., Hou, F., Amanai, K., and Jiang, J. 2005. Phosphorylation by double-time/CKIepsilon and CKIalpha targets cubitus interruptus for Slimb/beta-TRCP-mediated proteolytic processing. *Dev Cell* 9 (6):819-30.

Johnson, R. L., Laufer, E., Riddle, R. D., and Tabin, C. 1994. Ectopic expression of Sonic hedgehog alters dorsal-ventral patterning of somites. *Cell* 79 (7):1165-73.

Jordan, T., Hanson, I., Zaletayev, D., Hodgson, S., Prosser, J., Seawright, A., Hastie, N., and van Heyningen, V. 1992. The human PAX6 gene is mutated in two patients with aniridia. *Nat Genet* 1 (5):328-32.

Jung, H. S., Oropeza, V., and Thesleff, I. 1999. Shh, Bmp-2, Bmp-4 and Fgf-8 are associated with initiation and patterning of mouse tongue papillae. *Mech Dev* 81 (1-2):179-82.

Kaellen, B. 1965. Proliferation in the Embryonic Brain with Special Reference to the Overgrowth Phenomenon and Its Possible Relationship to Neoplasia. *Prog Brain Res* 14:263-78.

Karlsson, L., Bondjers, C., and Betsholtz, C. 1999. Roles for PDGF-A and sonic hedgehog in development of mesenchymal components of the hair follicle. *Development* 126 (12):2611-21.

Katoh, Y. and Katoh, M. 2004. Characterization of KIF7 gene in silico. *Int J Oncol* 25 (6):1881-6.

———. 2004. KIF27 is one of orthologs for Drosophila Costal-2. *Int J Oncol* 25 (6):1875-80.

Kaufman, R. M., Pham, C. T., and Ley, T. J. 1999. Transgenic analysis of a 100-kb human beta-globin cluster-containing DNA fragment propagated as a bacterial artificial chromosome. *Blood* 94 (9):3178-84.

Kiecker, C. and Niehrs, C. 2001. The role of prechordal mesendoderm in neural patterning. *Curr Opin Neurobiol* 11 (1):27-33.

Kleinjan, D. A., Seawright, A., Mella, S., Carr, C. B., Tyas, D. A., Simpson, T. I., Mason, J. O., Price, D. J., and van Heyningen, V. 2006. Long-range downstream enhancers are essential for Pax6 expression. *Dev Biol* 299 (2):563-81.

Kleinjan, D. A., Seawright, A., Schedl, A., Quinlan, R. A., Danes, S., and van Heyningen, V. 2001. Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6. *Hum Mol Genet* 10 (19):2049-59.

Kleinjan, D. J. and van Heyningen, V. 1998. Position effect in human genetic disease. *Hum Mol Genet* 7 (10):1611-8.

Kobayashi, H., Yamazaki, S., Takashima, S., Liu, W., Okuda, H., Yan, J., Fujii, Y., Hitomi, T., Harada, K. H., Habu, T., and Koizumi, A. 2013. Ablation of Rnf213 retards progression of diabetes in the Akita mouse. *Biochem Biophys Res Commun* 432 (3):519-25.

Kohtz, J. D., Baker, D. P., Corte, G., and Fishell, G. 1998. Regionalization within the mammalian telencephalon is mediated by changes in responsiveness to Sonic Hedgehog. *Development* 125 (24):5079-89.

Kokubu, C., Horie, K., Abe, K., Ikeda, R., Mizuno, S., Uno, Y., Ogiwara, S., Ohtsuka, M., Isotani, A., Okabe, M., Imai, K., and Takeda, J. 2009. A transposon-based chromosomal engineering method to survey a large cis-regulatory landscape in mice. *Nat Genet* 41 (8):946-52.

Komisarczuk, A. Z., Kawakami, K., and Becker, T. S. 2009. Cis-regulation and chromosomal rearrangement of the fgf8 locus after the teleost/tetrapod split. *Dev Biol* 336 (2):301-12.

Kontges, G. and Lumsden, A. 1996. Rhombencephalic neural crest segmentation is preserved throughout craniofacial ontogeny. *Development* 122 (10):3229-42.

Krauss, S., Concordet, J. P., and Ingham, P. W. 1993. A functionally conserved homolog of the Drosophila segment polarity gene hh is expressed in tissues with polarizing activity in zebrafish embryos. *Cell* 75 (7):1431-44.

Lang, M., Hadzhiev, Y., Siegel, N., Amemiya, C. T., Parada, C., Strahle, U., Becker, M. B., Muller, F., and Meyer, A. Conservation of shh cis-regulatory architecture of the coelacanth is consistent with its ancestral phylogenetic position. *Evodevo* 1 (1):11.

Le Douarin, N. M., Creuzet, S., Couly, G., and Dupin, E. 2004. Neural crest cell plasticity and its limits. *Development* 131 (19):4637-50.

Lee, J. J., von Kessler, D. P., Parks, S., and Beachy, P. A. 1992. Secretion and localized transcription suggest a role in positional signaling for products of the segmentation gene hedgehog. *Cell* 71 (1):33-50.

Lettice, L. A., Daniels, S., Sweeney, E., Venkataraman, S., Devenney, P. S., Gautier, P., Morrison, H., Fantes, J., Hill, R. E., and FitzPatrick, D. R. 2011. Enhancer-adoption as a mechanism of human developmental disease. *Hum Mutat* 32 (12):1492-9.

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E., and de Graaff, E. 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* 12 (14):1725-35.

Lettice, L. A. and Hill, R. E. 2005. Preaxial polydactyly: a model for defective long-range regulation in congenital abnormalities. *Curr Opin Genet Dev* 15 (3):294-300.

Lettice, L. A., Horikoshi, T., Heaney, S. J., van Baren, M. J., van der Linde, H. C., Breedveld, G. J., Joosse, M., Akarsu, N., Oostra, B. A., Endo, N., Shibata, M., Suzuki, M., Takahashi, E., Shinka, T., Nakahori, Y., Ayusawa, D., Nakabayashi, K., Scherer, S. W., Heutink, P., Hill, R. E., and Noji, S. 2002. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc Natl Acad Sci U S A* 99 (11):7548-53.

Lettice, L. A., Williamson, I., Devenney, P. S., Kilanowski, F., Dorin, J., and Hill, R. E. 2014. Development of five digits is controlled by a bipartite long-range cis-regulator. *Development* 141 (8):1715-25.

Lettice, L. A., Williamson, I., Wiltshire, J. H., Peluso, S., Devenney, P. S., Hill, A. E., Essafi, A., Hagman, J., Mort, R., Grimes, G., DeAngelis, C. L., and Hill, R. E. 2012. Opposing functions of the ETS factor family define Shh spatial expression in limb buds and underlie polydactyly. *Dev Cell* 22 (2):459-67.

Lewandoski, M., Sun, X., and Martin, G. R. 2000. Fgf8 signalling from the AER is essential for normal limb development. *Nat Genet* 26 (4):460-3.

Lewis, P. M., Dunn, M. P., McMahon, J. A., Logan, M., Martin, J. F., St-Jacques, B., and McMahon, A. P. 2001. Cholesterol modification of sonic hedgehog is required for long-range signaling activity and effective modulation of signaling by Ptc1. *Cell* 105 (5):599-612.

Li, B., Carey, M., and Workman, J. L. 2007. The role of chromatin during transcription. *Cell* 128 (4):707-19.

Li, C., Hu, L., Xiao, J., Chen, H., Li, J. T., Bellusci, S., Delanghe, S., and Minoo, P. 2005. Wnt5a regulates Shh and Fgf10 signaling during lung development. *Dev Biol* 287 (1):86-97.

Litingtung, Y., Dahn, R. D., Li, Y., Fallon, J. F., and Chiang, C. 2002. Shh and Gli3 are dispensable for limb skeleton formation but regulate digit number and identity. *Nature* 418 (6901):979-83.

Litingtung, Y., Lei, L., Westphal, H., and Chiang, C. 1998. Sonic hedgehog is essential to foregut development. *Nat Genet* 20 (1):58-61.

Liu, H. X., Ermilov, A., Grachtchouk, M., Li, L., Gumucio, D. L., Dlugosz, A. A., and Mistretta, C. M. 2013. Multiple Shh signaling centers participate in fungiform papilla and taste bud formation and maintenance. *Dev Biol* 382 (1):82-97.

Liu, P., Jenkins, N. A., and Copeland, N. G. 2003. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res* 13 (3):476-84.
Lohan, S., Spielmann, M., Doelken, S. C., Flottmann, R., Muhammad, F., Baig, S. M., Wajid, M., Hulsemann, W., Habenicht, R., Kjaer, K. W., Patil, S. J., Girisha, K. M., Abarca-Barriga, H. H., Mundlos, S., and Klopocki, E. 2014. Microduplications encompassing the Sonic hedgehog limb enhancer ZRS are associated with Haas-type polysyndactyly and Laurin-Sandrow syndrome. *Clin Genet*.
Loots, G. G. 2008. Genomic identification of regulatory elements by evolutionary sequence comparison and functional analysis. *Adv Genet* 61:269-93.
Maas, S. A. and Fallon, J. F. 2004. Isolation of the chicken Lmbr1 coding sequence and characterization of its role during chick limb development. *Dev Dyn* 229 (3):520-8.
Machold, R., Hayashi, S., Rutlin, M., Muzumdar, M. D., Nery, S., Corbin, J. G., Gritli-Linde, A., Dellovade, T., Porter, J. A., Rubin, L. L., Dudek, H., McMahon, A. P., and Fishell, G. 2003. Sonic hedgehog is required for progenitor cell maintenance in telencephalic stem cell niches. *Neuron* 39 (6):937-50.
Mansour, S., Hall, C. M., Pembrey, M. E., and Young, I. D. 1995. A clinical and genetic study of campomelic dysplasia. *J Med Genet* 32 (6):415-20.
Mansour, S., Offiah, A. C., McDowall, S., Sim, P., Tolmie, J., and Hall, C. 2002. The phenotype of survivors of campomelic dysplasia. *J Med Genet* 39 (8):597-602.
Mao, J., McGlinn, E., Huang, P., Tabin, C. J., and McMahon, A. P. 2009. Fgf-dependent Etv4/5 activity is required for posterior restriction of Sonic Hedgehog and promoting outgrowth of the vertebrate limb. *Dev Cell* 16 (4):600-6.
Marcucio, R. S., Cordero, D. R., Hu, D., and Helms, J. A. 2005. Molecular interactions coordinating the development of the forebrain and face. *Dev Biol* 284 (1):48-61.
Marigo, V., Davey, R. A., Zuo, Y., Cunningham, J. M., and Tabin, C. J. 1996. Biochemical evidence that patched is the Hedgehog receptor. *Nature* 384 (6605):176-9.
Marigo, V., Scott, M. P., Johnson, R. L., Goodrich, L. V., and Tabin, C. J. 1996. Conservation in hedgehog signaling: induction of a chicken patched homolog by Sonic hedgehog in the developing limb. *Development* 122 (4):1225-33.
Marigo, V. and Tabin, C. J. 1996. Regulation of patched by sonic hedgehog in the developing neural tube. *Proc Natl Acad Sci U S A* 93 (18):9346-51.
Marinic, M., Aktas, T., Ruf, S., and Spitz, F. 2013. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev Cell* 24 (5):530-42.
Mates, L., Chuah, M. K., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D. P., Schmitt, A., Becker, K., Matrai, J., Ma, L., Samara-Kuko, E., Gysemans, C., Pryputniewicz, D., Miskey, C., Fletcher, B., VandenDriessche, T., Ivics, Z., and Izsvak, Z. 2009. Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat Genet* 41 (6):753-61.
Matessi, C. and Schneider, K. A. 2009. Optimization under frequency-dependent selection. *Theor Popul Biol* 76 (1):1-12.
McGaughey, D. M., Vinton, R. M., Huynh, J., Al-Saif, A., Beer, M. A., and McCallion, A. S. 2008. Metrics of sequence constraint overlook regulatory sequences in an exhaustive analysis at phox2b. *Genome Res* 18 (2):252-60.
Melton, D. W. 2002. Gene-targeting strategies. *Methods Mol Biol* 180:151-73.
Meulemans, D. and Bronner-Fraser, M. 2004. Gene-regulatory interactions in neural crest evolution and development. *Dev Cell* 7 (3):291-9.
Meyers, E. N., Lewandoski, M., and Martin, G. R. 1998. An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination. *Nat Genet* 18 (2):136-41.
Mill, P., Mo, R., Fu, H., Grachtchouk, M., Kim, P. C., Dlugosz, A. A., and Hui, C. C. 2003. Sonic hedgehog-dependent activation of Gli2 is essential for embryonic hair follicle development. *Genes Dev* 17 (2):282-94.

Miller, L. A., Wert, S. E., and Whitsett, J. A. 2001. Immunolocalization of sonic hedgehog (Shh) in developing mouse lung. *J Histochem Cytochem* 49 (12):1593-604.
Montavon, T. and Duboule, D. 2013. Chromatin organization and global regulation of Hox gene clusters. *Philos Trans R Soc Lond B Biol Sci* 368 (1620):20120367.
Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F., and Duboule, D. 2011. A regulatory archipelago controls Hox genes transcription in digits. *Cell* 147 (5):1132-45.
Moon, A. M. and Capecchi, M. R. 2000. Fgf8 is required for outgrowth and patterning of the limbs. *Nat Genet* 26 (4):455-9.
Mullor, J. L., Calleja, M., Capdevila, J., and Guerrero, I. 1997. Hedgehog activity, independent of decapentaplegic, participates in wing disc patterning. *Development* 124 (6):1227-37.
Myers, B. R., Sever, N., Chong, Y. C., Kim, J., Belani, J. D., Rychnovsky, S., Bazan, J. F., and Beachy, P. A. 2013. Hedgehog pathway modulation by multiple lipid binding sites on the smoothened effector of signal response. *Dev Cell* 26 (4):346-57.
Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W., and Roder, J. C. 1993. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc Natl Acad Sci U S A* 90 (18):8424-8.
Ng, L. J., Wheatley, S., Muscat, G. E., Conway-Campbell, J., Bowles, J., Wright, E., Bell, D. M., Tam, P. P., Cheah, K. S., and Koopman, P. 1997. SOX9 binds DNA, activates transcription, and coexpresses with type II collagen during chondrogenesis in the mouse. *Dev Biol* 183 (1):108-21.
Niedermaier, M., Schwabe, G. C., Fees, S., Helmrich, A., Brieske, N., Seemann, P., Hecht, J., Seitz, V., Stricker, S., Leschik, G., Schrock, E., Selby, P. B., and Mundlos, S. 2005. An inversion involving the mouse Shh locus results in brachydactyly through dysregulation of Shh expression. *J Clin Invest* 115 (4):900-9.
Nielsen, L. B., McCormick, S. P., Pierotti, V., Tam, C., Gunn, M. D., Shizuya, H., and Young, S. G. 1997. Human apolipoprotein B transgenic mice generated with 207- and 145-kilobase pair bacterial artificial chromosomes. Evidence that a distant 5'-element confers appropriate transgene expression in the intestine. *J Biol Chem* 272 (47):29752-8.
Nobrega, M. A., Zhu, Y., Plajzer-Frick, I., Afzal, V., and Rubin, E. M. 2004. Megabase deletions of gene deserts result in viable mice. *Nature* 431 (7011):988-93.
Nusse, R. 2003. Wnts and Hedgehogs: lipid-modified proteins and similarities in signaling mechanisms at the cell surface. *Development* 130 (22):5297-305.
Nusslein-Volhard, C. and Wieschaus, E. 1980. Mutations affecting segment number and polarity in Drosophila. *Nature* 287 (5785):795-801.
Ohkubo, Y., Chiang, C., and Rubenstein, J. L. 2002. Coordinate regulation and synergistic actions of BMP4, SHH and FGF8 in the rostral prosencephalon regulate morphogenesis of the telencephalic and optic vesicles. *Neuroscience* 111 (1):1-17.
Ohtsuki, S., Levine, M., and Cai, H. N. 1998. Different core promoters possess distinct regulatory activities in the Drosophila embryo. *Genes Dev* 12 (4):547-56.
Olson, E. N., Arnold, H. H., Rigby, P. W., and Wold, B. J. 1996. Know your neighbors: three phenotypes in null mutants of the myogenic bHLH gene MRF4. *Cell* 85 (1):1-4.
Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J. A., Lopes, S., Reik, W., and Fraser, P. 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* 36 (10):1065-71.
Osoegawa, K., Tateno, M., Woon, P. Y., Frengen, E., Mammoser, A. G., Catanese, J. J., Hayashizaki, Y., and de Jong, P. J. 2000. Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res* 10 (1):116-28.

Osumi-Yamashita, N., Ninomiya, Y., Doi, H., and Eto, K. 1994. The contribution of both forebrain and midbrain crest cells to the mesenchyme in the frontonasal mass of mouse embryos. *Dev Biol* 164 (2):409-19.

Ovcharenko, I., Loots, G. G., Nobrega, M. A., Hardison, R. C., Miller, W., and Stubbs, L. 2005. Evolution and functional classification of vertebrate gene deserts. *Genome Res* 15 (1):137-45.

Pabst, O., Herbrand, H., Takuma, N., and Arnold, H. H. 2000. NKX2 gene expression in neuroectoderm but not in mesendodermally derived structures depends on sonic hedgehog in mouse embryos. *Dev Genes Evol* 210 (1):47-50.

Paek, H., Antoine, M. W., Diaz, F., and Hebert, J. M. 2012. Increased beta-catenin activity in the anterior neural plate induces ectopic mid-hindbrain characteristics. *Dev Dyn* 241 (2):242-6.

Panman, L., Drenth, T., Tewelscher, P., Zuniga, A., and Zeller, R. 2005. Genetic interaction of Gli3 and Alx4 during limb development. *Int J Dev Biol* 49 (4):443-8.

Papaioannou, V. E. and Behringer, R. R. 2012. Early embryonic lethality in genetically engineered mice: diagnosis and phenotypic analysis. *Vet Pathol* 49 (1):64-70.

Peng, Y. C., Levine, C. M., Zahid, S., Wilson, E. L., and Joyner, A. L. 2013. Sonic hedgehog signals to multiple prostate stromal stem cells that replenish distinct stromal subtypes during regeneration. *Proc Natl Acad Sci U S A* 110 (51):20611-6.

Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K. D., Plajzer-Frick, I., Akiyama, J., De Val, S., Afzal, V., Black, B. L., Couronne, O., Eisen, M. B., Visel, A., and Rubin, E. M. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444 (7118):499-502.

Pepicelli, C. V., Lewis, P. M., and McMahon, A. P. 1998. Sonic hedgehog regulates branching morphogenesis in the mammalian lung. *Curr Biol* 8 (19):1083-6.

Perry, M. W., Boettiger, A. N., Bothma, J. P., and Levine, M. 2010. Shadow enhancers foster robustness of Drosophila gastrulation. *Curr Biol* 20 (17):1562-7.

Phillips, J. E. and Corces, V. G. 2009. CTCF: master weaver of the genome. *Cell* 137 (7):1194-211.

Placzek, M., Dodd, J., and Jessell, T. M. 2000. Discussion point. The case for floor plate induction by the notochord. *Curr Opin Neurobiol* 10 (1):15-22.

Qin, Y., Kong, L. K., Poirier, C., Truong, C., Overbeek, P. A., and Bishop, C. E. 2004. Long-range activation of Sox9 in Odd Sex (Ods) mice. *Hum Mol Genet* 13 (12):1213-8.

Quinlan, R. J., Tobin, J. L., and Beales, P. L. 2008. Modeling ciliopathies: Primary cilia in development and disease. *Curr Top Dev Biol* 84:249-310.

Rada-Iglesias, A. and Wysocka, J. 2011. Epigenomics of human embryonic stem cells and induced pluripotent stem cells: insights into pluripotency and implications for disease. *Genome Med* 3 (6):36.

Rallu, M., Machold, R., Gaiano, N., Corbin, J. G., McMahon, A. P., and Fishell, G. 2002. Dorsoventral patterning is established in the telencephalon of mutants lacking both Gli3 and Hedgehog signaling. *Development* 129 (21):4963-74.

Ramalho-Santos, M., Melton, D. A., and McMahon, A. P. 2000. Hedgehog signals regulate multiple aspects of gastrointestinal development. *Development* 127 (12):2763-72.

Riddle, R. D., Johnson, R. L., Laufer, E., and Tabin, C. 1993. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* 75 (7):1401-16.

Roberts, D. J., Johnson, R. L., Burke, A. C., Nelson, C. E., Morgan, B. A., and Tabin, C. 1995. Sonic hedgehog is an endodermal signal inducing Bmp-4 and Hox genes during induction and regionalization of the chick hindgut. *Development* 121 (10):3163-74.

Rock, J. R., Lopez, M. C., Baker, H. V., and Harfe, B. D. 2007. Identification of genes expressed in the mouse limb using a novel ZPA microarray approach. *Gene Expr Patterns* 8 (1):19-26.

Roelink, H., Porter, J. A., Chiang, C., Tanabe, Y., Chang, D. T., Beachy, P. A., and Jessell, T. M. 1995. Floor plate and motor neuron induction by different concentrations of the amino-terminal cleavage product of sonic hedgehog autoproteolysis. *Cell* 81 (3):445-55.

Roessler, E., Belloni, E., Gaudenz, K., Jay, P., Berta, P., Scherer, S. W., Tsui, L. C., and Muenke, M. 1996. Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nat Genet* 14 (3):357-60.

Roessler, E., Ward, D. E., Gaudenz, K., Belloni, E., Scherer, S. W., Donnai, D., Siegel-Bartelt, J., Tsui, L. C., and Muenke, M. 1997. Cytogenetic rearrangements involving the loss of the Sonic Hedgehog gene at 7q36 cause holoprosencephaly. *Hum Genet* 100 (2):172-81.

Rohatgi, R., Milenkovic, L., and Scott, M. P. 2007. Patched1 regulates hedgehog signaling at the primary cilium. *Science* 317 (5836):372-6.

Rosenbaum, J. L. and Witman, G. B. 2002. Intraflagellar transport. *Nat Rev Mol Cell Biol* 3 (11):813-25.

Ruel, L., Rodriguez, R., Gallet, A., Lavenant-Staccini, L., and Therond, P. P. 2003. Stability and association of Smoothened, Costal2 and Fused with Cubitus interruptus are regulated by Hedgehog. *Nat Cell Biol* 5 (10):907-13.

Ruf, S., Symmons, O., Uslu, V. V., Dolle, D., Hot, C., Ettwiller, L., and Spitz, F. 2011. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat Genet* 43 (4):379-86.

Ruiz i Altaba, A., Palma, V., and Dahmane, N. 2002. Hedgehog-Gli signalling and the growth of the brain. *Nat Rev Neurosci* 3 (1):24-33.

Sagai, T., Amano, T., Tamura, M., Mizushina, Y., Sumiyama, K., and Shiroishi, T. 2009. A cluster of three long-range enhancers directs regional Shh expression in the epithelial linings. *Development* 136 (10):1665-74.

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., and Shiroishi, T. 2005. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* 132 (4):797-803.

Sagai, T., Masuya, H., Tamura, M., Shimizu, K., Yada, Y., Wakana, S., Gondo, Y., Noda, T., and Shiroishi, T. 2004. Phylogenetic conservation of a limb-specific, cis-acting regulator of Sonic hedgehog (Shh). *Mamm Genome* 15 (1):23-34.

Sander, J. D. and Joung, J. K. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* 32 (4):347-55.

Santagati, F. and Rijli, F. M. 2003. Cranial neural crest and the building of the vertebrate head. *Nat Rev Neurosci* 4 (10):806-18.

Sarpal, R., Todi, S. V., Sivan-Loukianova, E., Shirolikar, S., Subramanian, N., Raff, E. C., Erickson, J. W., Ray, K., and Eberl, D. F. 2003. Drosophila KAP interacts with the kinesin II motor subunit KLP64D to assemble chordotonal sensory cilia, but not sperm tails. *Curr Biol* 13 (19):1687-96.

Schneider, R. A. 1999. Neural crest can form cartilages normally derived from mesoderm during development of the avian head skeleton. *Dev Biol* 208 (2):441-55.

Schneider, R. A., Hu, D., Rubenstein, J. L., Maden, M., and Helms, J. A. 2001. Local retinoid signaling coordinates forebrain and facial morphogenesis by maintaining FGF8 and SHH. *Development* 128 (14):2755-67.

Schuurmans, C. and Guillemot, F. 2002. Molecular mechanisms underlying cell fate specification in the developing telencephalon. *Curr Opin Neurobiol* 12 (1):26-34.

Sharpe, J., Ahlgren, U., Perry, P., Hill, B., Ross, A., Hecksher-Sorensen, J., Baldock, R., and Davidson, D. 2002. Optical projection tomography as a tool for 3D microscopy and gene expression studies. *Science* 296 (5567):541-5.

Sharpe, J., Lettice, L., Hecksher-Sorensen, J., Fox, M., Hill, R., and Krumlauf, R. 1999. Identification of sonic hedgehog as a candidate gene responsible for the polydactylous mouse mutant Sasquatch. *Curr Biol* 9 (2):97-100.

Sharpe, J., Nonchev, S., Gould, A., Whiting, J., and Krumlauf, R. 1998. Selectivity, sharing and competitive interactions in the regulation of Hoxb genes. *EMBO J* 17 (6):1788-98.

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V. V., and Ren, B. 2012. A map of the cis-regulatory sequences in the mouse genome. *Nature* 488 (7409):116-20.

Smallwood, A. and Ren, B. 2013. Genome organization and long-range regulation of gene expression by enhancers. *Curr Opin Cell Biol* 25 (3):387-94.

Soshnikova, N., Montavon, T., Leleu, M., Galjart, N., and Duboule, D. 2010. Functional analysis of CTCF during mammalian limb development. *Dev Cell* 19 (6):819-30.

Spitz, F. and Duboule, D. 2008. Global control regions and regulatory landscapes in vertebrate development and evolution. *Adv Genet* 61:175-205.

Spitz, F., Gonzalez, F., and Duboule, D. 2003. A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* 113 (3):405-17.

Spitz, F., Herkenne, C., Morris, M. A., and Duboule, D. 2005. Inversion-induced disruption of the Hoxd cluster leads to the partition of regulatory landscapes. *Nat Genet* 37 (8):889-93.

St-Jacques, B., Dassule, H. R., Karavanova, I., Botchkarev, V. A., Li, J., Danielian, P. S., McMahon, J. A., Lewis, P. M., Paus, R., and McMahon, A. P. 1998. Sonic hedgehog signaling is essential for hair development. *Curr Biol* 8 (19):1058-68.

Stone, D. M., Hynes, M., Armanini, M., Swanson, T. A., Gu, Q., Johnson, R. L., Scott, M. P., Pennica, D., Goddard, A., Phillips, H., Noll, M., Hooper, J. E., de Sauvage, F., and Rosenthal, A. 1996. The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* 384 (6605):129-34.

Struhl, G., Barbash, D. A., and Lawrence, P. A. 1997. Hedgehog acts by distinct gradient and signal relay mechanisms to organise cell type and cell polarity in the Drosophila abdomen. *Development* 124 (11):2155-65.

Sun, M., Ma, F., Zeng, X., Liu, Q., Zhao, X. L., Wu, F. X., Wu, G. P., Zhang, Z. F., Gu, B., Zhao, Y. F., Tian, S. H., Lin, B., Kong, X. Y., Zhang, X. L., Yang, W., Lo, W. H., and Zhang, X. 2008. Triphalangeal thumb-polysyndactyly syndrome and syndactyly type IV are caused by genomic duplications involving the long range, limb-specific SHH enhancer. *J Med Genet* 45 (9):589-95.

Sun, X., Mariani, F. V., and Martin, G. R. 2002. Functions of FGF signalling from the apical ectodermal ridge in limb development. *Nature* 418 (6897):501-8.

Sun, X., Meyers, E. N., Lewandoski, M., and Martin, G. R. 1999. Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. *Genes Dev* 13 (14):1834-46.

Symmons, O. and Spitz, F. 2013. From remote enhancers to gene regulation: charting the genome's regulatory landscapes. *Philos Trans R Soc Lond B Biol Sci* 368 (1620):20120358.

Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. 2014. Functional and topological characteristics of mammalian regulatory domains. *Genome Res* 24 (3):390-400.

Tarchini, B. and Duboule, D. 2006. Control of Hoxd genes' collinearity during early limb development. *Dev Cell* 10 (1):93-103.

te Welscher, P., Fernandez-Teran, M., Ros, M. A., and Zeller, R. 2002. Mutual genetic antagonism involving GLI3 and dHAND prepatterns the vertebrate limb bud mesenchyme prior to SHH signaling. *Genes Dev* 16 (4):421-6.

Teillet, M. A., Lapointe, F., and Le Douarin, N. M. 1998. The relationships between notochord and floor plate in vertebrate development revisited. *Proc Natl Acad Sci U S A* 95 (20):11733-8.

Ten Have-Opbroek, A. A. 1991. Lung development in the mouse embryo. *Exp Lung Res* 17 (2):111-30.

Tenzen, T., Allen, B. L., Cole, F., Kang, J. S., Krauss, R. S., and McMahon, A. P. 2006. The cell surface membrane proteins Cdo and Boc are components and targets of the Hedgehog signaling pathway and feedback network in mice. *Dev Cell* 10 (5):647-56.

Thaler, J., Harrison, K., Sharma, K., Lettieri, K., Kehrl, J., and Pfaff, S. L. 1999. Active suppression of interneuron programs within developing motor neurons revealed by analysis of homeodomain factor HB9. *Neuron* 23 (4):675-87.

Theil, T., Alvarez-Bolado, G., Walter, A., and Ruther, U. 1999. Gli3 is required for Emx gene expression during dorsal telencephalon development. *Development* 126 (16):3561-71.

Tickle, C., Summerbell, D., and Wolpert, L. 1975. Positional signalling and specification of digits in chick limb morphogenesis. *Nature* 254 (5497):199-202.

Trumpp, A., Blundell, P. A., de la Pompa, J. L., and Zeller, R. 1992. The chicken limb deformity gene encodes nuclear proteins expressed in specific cell types during morphogenesis. *Genes Dev* 6 (1):14-28.

Tsukiji, N., Amano, T., and Shiroishi, T. 2013. A novel regulatory element for Shh expression in the lung and gut of mouse embryos. *Mech Dev*.

Valancius, V. and Smithies, O. 1991. Double-strand gap repair in a mammalian gene targeting reaction. *Mol Cell Biol* 11 (9):4389-97.

van Baren, M. J., van der Linde, H. C., Breedveld, G. J., Baarends, W. M., Rizzu, P., de Graaff, E., Oostra, B. A., and Heutink, P. 2002. A double RING-H2 domain in RNF32, a gene expressed during sperm formation. *Biochem Biophys Res Commun* 292 (1):58-65.

Varjosalo, M., Li, S. P., and Taipale, J. 2006. Divergence of hedgehog signal transduction mechanism between Drosophila and mammals. *Dev Cell* 10 (2):177-86.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E. M., and Pennacchio, L. A. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457 (7231):854-8.

Visel, A., Rubin, E. M., and Pennacchio, L. A. 2009. Genomic views of distant-acting enhancers. *Nature* 461 (7261):199-205.

Vortkamp, A., Lee, K., Lanske, B., Segre, G. V., Kronenberg, H. M., and Tabin, C. J. 1996. Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. *Science* 273 (5275):613-22.

Wagner, T., Wirth, J., Meyer, J., Zabel, B., Held, M., Zimmer, J., Pasantes, J., Bricarelli, F. D., Keutel, J., Hustert, E., Wolf, U., Tommerup, N., Schempp, W., and Scherer, G. 1994. Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. *Cell* 79 (6):1111-20.

Wallace, J. A. and Felsenfeld, G. 2007. We gather together: insulators and genome organization. *Curr Opin Genet Dev* 17 (5):400-7.

Wang, C., Ruther, U., and Wang, B. 2007. The Shh-independent activator function of the full-length Gli3 protein and its role in vertebrate limb digit patterning. *Dev Biol* 305 (2):460-9.

Wang, L. C., Liu, Z. Y., Gambardella, L., Delacour, A., Shapiro, R., Yang, J., Sizing, I., Rayhorn, P., Garber, E. A., Benjamin, C. D., Williams, K. P., Taylor, F. R., Barrandon, Y., Ling, L., and Burkly, L. C. 2000. Regular articles: conditional disruption of hedgehog signaling pathway defines its critical role in hair development and regeneration. *J Invest Dermatol* 114 (5):901-8.

Ward, J. M., Elmore, S. A., and Foley, J. F. 2012. Pathology methods for the evaluation of embryonic and perinatal developmental defects and lethality in genetically engineered mice. *Vet Pathol* 49 (1):71-84.

Wieczorek, D., Pawlik, B., Li, Y., Akarsu, N. A., Caliebe, A., May, K. J., Schweiger, B., Vargas, F. R., Balci, S., Gillessen-Kaesbach, G., and Wollnik, B. 2010. A specific mutation in the distant sonic hedgehog (SHH) cis-regulator (ZRS) causes Werner mesomelic syndrome (WMS) while complete ZRS duplications underlie Haas type polysyndactyly and preaxial polydactyly (PPD) with or without triphalangeal thumb. *Hum Mutat* 31 (1):81-9.

Wijgerde, M., McMahon, J. A., Rule, M., and McMahon, A. P. 2002. A direct requirement for Hedgehog signaling for normal specification of all ventral progenitor domains in the presumptive mammalian spinal cord. *Genes Dev* 16 (22):2849-64.

Wilson, S. W. and Houart, C. 2004. Early steps in the development of the forebrain. *Dev Cell* 6 (2):167-81.

Wolpert, L. 1969. Positional information and the spatial pattern of cellular differentiation. *J Theor Biol* 25 (1):1-47.

Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., Smith, S. F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W., Edwards, Y. J., Cooke, J. E., and Elgar, G. 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol* 3 (1):e7.

Wunderle, V. M., Critcher, R., Hastie, N., Goodfellow, P. N., and Schedl, A. 1998. Deletion of long-range regulatory elements upstream of SOX9 causes campomelic dysplasia. *Proc Natl Acad Sci U S A* 95 (18):10649-54.

Xu, Q., Wonders, C. P., and Anderson, S. A. 2005. Sonic hedgehog maintains the identity of cortical interneuron progenitors in the ventral telencephalon. *Development* 132 (22):4987-98.

Yamagishi, C., Yamagishi, H., Maeda, J., Tsuchihashi, T., Ivey, K., Hu, T., and Srivastava, D. 2006. Sonic hedgehog is essential for first pharyngeal arch development. *Pediatr Res* 59 (3):349-54.

Yang, G. S., Banks, K. G., Bonaguro, R. J., Wilson, G., Dreolini, L., de Leeuw, C. N., Liu, L., Swanson, D. J., Goldowitz, D., Holt, R. A., and Simpson, E. M. 2009. Next generation tools for high-throughput promoter and expression analysis employing single-copy knock-ins at the Hprt1 locus. *Genomics* 93 (3):196-204.

Ye, W., Shimamura, K., Rubenstein, J. L., Hynes, M. A., and Rosenthal, A. 1998. FGF and Shh signals control dopaminergic and serotonergic cell fate in the anterior neural plate. *Cell* 93 (5):755-66.

Zakany, J. and Duboule, D. 2007. The role of Hox genes during vertebrate limb development. *Curr Opin Genet Dev* 17 (4):359-66.

Zaki, P. A., Quinn, J. C., and Price, D. J. 2003. Mouse models of telencephalic development. *Curr Opin Genet Dev* 13 (4):423-37.

Zeller, R. and Zuniga, A. 2007. Shh and Gremlin1 chromosomal landscapes in development and disease. *Curr Opin Genet Dev* 17 (5):428-34.

Zhang, Q., Zhang, L., Wang, B., Ou, C. Y., Chien, C. T., and Jiang, J. 2006. A hedgehog-induced BTB protein modulates hedgehog signaling by degrading Ci/Gli transcription factor. *Dev Cell* 10 (6):719-29.

Zhang, X. M., Ramalho-Santos, M., and McMahon, A. P. 2001. Smoothened mutants reveal redundant roles for Shh and Ihh signaling including regulation of L/R asymmetry by the mouse node. *Cell* 105 (6):781-92.

———. 2001. Smoothened mutants reveal redundant roles for Shh and Ihh signaling including regulation of L/R symmetry by the mouse node. *Cell* 106 (2):781-92.

Zhang, Z., Verheyden, J. M., Hassell, J. A., and Sun, X. 2009. FGF-regulated Etv genes are essential for repressing Shh expression in mouse limb buds. *Dev Cell* 16 (4):607-13.

Zhu, J., Nakamura, E., Nguyen, M. T., Bao, X., Akiyama, H., and Mackem, S. 2008. Uncoupling Sonic hedgehog control of pattern and expansion of the developing limb bud. *Dev Cell* 14 (4):624-32.

Zuniga, A., Michos, O., Spitz, F., Haramis, A. P., Panman, L., Galli, A., Vintersten, K., Klasen, C., Mansfield, W., Kuc, S., Duboule, D., Dono, R., and Zeller, R. 2004. Mouse limb deformity mutations disrupt a global control region within the large regulatory landscape required for Gremlin expression. *Genes Dev* 18 (13):1553-64.